A COURSE OF HIGHER MATHEMATICS

V. I. Smirnov

Volume I

Elementary Calculus

ADIWES INTERNATIONAL SERIES
IN MATHEMATICS

A. J. LOHWATER
Consulting Editor

A COURSE OF
Higher Mathematics

VOLUME I

V. I. SMIRNOV

Translated by

D. E. BROWN

Translation edited and additions made by

I. N. SNEDDON
Simson Professor in Mathematics
University of Glasgow

PERGAMON PRESS
OXFORD · LONDON · EDINBURGH · NEW YORK
PARIS · FRANKFURT

ADDISON-WESLEY PUBLISHING COMPANY, INC.
READING, MASSACHUSETTS · PALO ALTO · LONDON
1964

U. S. A. Edition distributed by
ADDISON-WESLEY PUBLISHING COMPANY, INC.
Reading, Massachusetts • Palo Alto • London

PERGAMON PRESS
International Series of Monographs in
PURE AND APPLIED MATHEMATICS

Volume 57

Library of Congress Catalog Card No. 63-10134

This translation has been made from
the Sixteenth (revised) Russian Edition of
V. I. Smirnov's book
Курс высшей математики (Kurs vysshei matematiki),
published in 1957 by Fizmatgiz, Moscow
MADE IN GREAT BRITAIN

INTRODUCTION

THIS is the first volume of a five-volume course of higher mathematics which has been studied by Soviet mathematicians, physicists and engineers for forty years. In the first two editions (1924, 1927), which were practically identical, this first volume was written jointly by J. D. Tamarkin and V. I. Smirnov, but on the title page of later editions, prepared without the late Professor Tamarkin's cognizance and deviating from the two earlier editions in many respects, Professor Smirnov's name appears alone.

Professor Tamarkin's career and his contributions to both Russian and American mathematics are well known to British and American readers, but the achievements of Professor Smirnov are known to a more restricted circle. Vladimir Ivanovitch Smirnov, who was born in 1887, has had a distinguished career in research and teaching which fits him ideally for the writing of a comprehensive work of extensive proportions. His research has been mainly in the theory of functions and of differential equations but he has made valuable contributions to applied mathematics and, in particular, to theoretical seismology and all his work has been characterized by a broad scientific outlook and he has done more than any other Soviet mathematician to maintain and strengthen the connections between mathematics and physics. His pupils, among whom are numbered S. L. Sobolev, N. E. Kochin and I. A. Lappo-Danilevskii, have maintained this tradition of working in both pure and applied mathematics, a tradition which Smirnov inherited from his teacher V. A. Steklov.

Professor Smirnov's teaching experience in the old Institute of Transport, in a technical high school, in the Physics Department of the Mathematics and Physics Faculty of the University of Leningrad, and as Director of the Theoretical Section of the Institute of Seismology, Moscow, led him to study the design of a special course of higher mathematics for physicists and engineers, a project in the course of which he received the counsel of his many physicist friends particularly V. A. Fock and T. V. Kravets. The five-volume set of which the present volume is the first is the outcome of that study. It is, of course, designed as a first course for pure mathematicians in the

topics considered as well as for students and research workers whose main interest lies in the applications of mathematics.

The whole work is notable not only for the wealth of the illustrations it draws from physics and technology to illuminate points in pure mathematics, but also for the clarity of the exposition. This has already been recognized in the Soviet Union by the esteem by which the author's work is held by academic teachers, by the award in 1947 of the State Prize (previously called the Stalin Prize) to the author for this work, and it is to be hoped that through Mr. Brown's translation its merits will become just as well known in the English-speaking world.

The present volume is an introduction to calculus and to the principles of mathematical analysis including some introductory material on functions of several variables as well as on functions of a single variable. As well as providing the material necessary for the understanding of the methods of mathematical physics it is an excellent introduction to these subjects for students of pure mathematics.

I. N. SNEDDON

PREFACE TO THE EIGHTH RUSSIAN EDITION

THE present edition differs very considerably from the last. The material relating to analytic geometry has been excluded, and the remaining material has been rearranged as a result. In particular, applications of the differential calculus to geometry are now to be found collected in § 7 (Chapter II). A chapter has been added which was previously the first of Volume II, dealing with complex numbers, the basic properties of integral polynomials, and the systematic integration of functions.

Further substantial additions must be mentioned, apart from the various minor additions and modifications to the text. In view of the fact that quite subtle and difficult problems of higher analysis are encountered in later volumes, it was thought useful to give the theory of irrational numbers, and its use in proving tests for the existence of limits and the properties of continuous functions, at the end of § 2 (Chapter I) after the theory of limits. A rigorous definition and study of the properties of the elementary function is also to be found there. The proof of the existence of implicit functions is included in Chapter V, dealing with functions of several variables.

The text is arranged so that the large type can be read independently. The small type sections contain examples, some additional particular problems, all the theoretical material referred to above, and the final section of Chapter IV, which deals with theory of a more difficult kind.

My sincere thanks are due to Professor G. M. Fikhtengol'ts for a number of valuable suggestions regarding the text, which I have incorporated during the final revision of the book.

PREFACE TO THE SIXTEENTH RUSSIAN EDITION

THE basic text and plan of the book have remained unchanged in the present edition, though there are a number of alterations due to the requirements of accuracy and completeness. This refers especially to applications of the differential and integral calculus to geometry.

V. SMIRNOV

PREFACE TO THE EIGHTH RUSSIAN EDITION

THE present edition differs very considerably from the last. The material relating to analytic geometry has been excluded, and the remaining material has been rearranged as a result. In particular, applications of the differential calculus to geometry are now to be found collected in § 7 (Chapter II). A chapter has been added which was previously the first of Volume II, dealing with complex numbers, the basic properties of integral polynomials, and the systematic integration of functions.

Further substantial additions must be mentioned, apart from the various minor additions and modifications to the text. In view of the fact that quite subtle and difficult problems of higher analysis are encountered in later volumes, it was thought useful to give the theory of irrational numbers, and its use in proving tests for the existence of limits and the properties of continuous functions, at the end of § 2 (Chapter I) after the theory of limits. A rigorous definition and study of the properties of the elementary function is also to be found there. The proof of the existence of implicit functions is included in Chapter V, dealing with functions of several variables.

The text is arranged so that the large type can be read independently. The small type sections contain examples, some additional particular problems, all the theoretical material referred to above, and the final section of Chapter IV, which deals with theory of a more difficult kind.

My sincere thanks are due to Professor G. M. Fikhtengol'ts for a number of valuable suggestions regarding the text, which I have incorporated during the final revision of the book.

PREFACE TO THE SIXTEENTH RUSSIAN EDITION

THE basic text and plan of the book have remained unchanged in the present edition, though there are a number of alterations due to the requirements of accuracy and completeness. This refers especially to applications of the differential and integral calculus to geometry.

V. SMIRNOV

CHAPTER I

FUNCTIONAL RELATIONSHIPS AND THE THEORY OF LIMITS

§ 1. Variables

1. Magnitude and its measurement. Mathematical analysis has a fundamental importance for exact science; unlike the other sciences, each of which has an interest only in some limited aspect of the world around us, mathematics is concerned with the most general properties inherent in all phenomena that are open to scientific investigation.

One of the fundamental concepts is that of *magnitude and its measurement*. It is characteristic of a magnitude that it can be measured, i.e. it can be compared in one way or another with some specific magnitude of the sort which is accepted as *the unit of measurement*. The process of comparison itself depends on the nature of the magnitude in question and is called measurement. Measurement results in an *abstract number* being obtained, expressing the ratio of the observed magnitude to the magnitude accepted as the unit of measurement.

Every law of nature gives us a correlation between magnitudes, or more exactly, between numbers expressing these magnitudes. It is precisely the object of mathematics to study numbers and the various correlations between them, independently of the concrete nature of the magnitudes and laws which lead us to these numbers and correlations.

Thus, *every magnitude is related by its measurement to an abstract number*. This number depends essentially, however, on the unit assumed for the measurement, or on the *scale*. On increasing this unit, the number measuring a given magnitude decreases, and conversely, the number increases on decreasing the unit.

The choice of scale is governed by the character of the magnitude concerned and by the circumstances in which the measurements are carried out. The size of the scale used for measuring one and the same

magnitude can vary within the widest possible limits — for instance, in measuring length in accurate optical studies the accepted unit of length is an Angstrom (one ten-millionth of a millimetre, 10^{-10} m); whereas use is made in astronomy of a unit of length called a *light-year*, i.e. the distance travelled by light in the course of a year (light travels approximately 300,000 km in one second).

2. Number. The number which is obtained as a result of measurement may be *integral* (if the unit goes an integral number of times into the magnitude concerned), *fractional* (if another unit exists, which goes an integral number of times both into the measured magnitude and into the unit previously chosen — or in short, when the measured magnitude is *commensurable* with the unit of measurement) and finally, *irrational* (when no such common measure exists, i.e. the given magnitude proves *incommensurable* with the unit of measurement).

It is shown in elementary geometry, for instance, that the diagonal of a square is incommensurable with its side, so that, if we measure the diagonal of a square using the length of side as unit, the number $\sqrt{2}$ obtained by measurement is irrational. The number π is similarly irrational, obtained on measuring the circumference of a circle, the diameter of which is taken as unit.

Reference can usefully be made to decimal fractions, in order to understand the idea of irrational numbers. As is known from arithmetic, every rational number can be represented in the form of either a finite or an infinite decimal fraction, the infinite fraction being periodic in the latter case (simple periodic or compound periodic). For instance, on carrying out division of the numerator by the denominator in accordance with the rule for division into decimal fractions, we obtain:

$$\frac{5}{33} = 0.151515\ldots = 0.1(5),$$

$$\frac{5}{18} = 0.2777\ldots = 0.2\,(7).$$

Conversely, as is known from arithmetic, every periodic decimal fraction expresses a rational number.

In measuring a magnitude, incommensurable with the unit taken, we can first reckon how many times a full unit goes into the measured magnitude, then how many times a tenth of a unit goes into the remainder obtained, then how many times a hundredth of a unit goes into the new remainder and so on. Measurement of a magnitude, incommen-

surable with the unit, will thus lead to the formation of an infinite non-periodic decimal fraction. An infinite fraction of this sort corresponds to every irrational number, and conversely, to every infinite non-periodic decimal fraction there corresponds a certain irrational number. If only a few of the first decimal places are retained in this infinite decimal fraction, an approximate value is obtained below the irrational number represented by this fraction. Thus, for example, on extracting the square root in accordance with the usual rule to the third decimal place, we obtain:

$$\sqrt{2} = 1.414\ldots$$

The numbers 1.414 and 1.415 are approximate values of $\sqrt{2}$ with an accuracy of one-thousandth, below and above.

Decimal places can be used for comparing the sizes of irrational numbers with each other, and with rational numbers.

Magnitudes of different signs, positive and negative, have to be considered in many cases (temperatures above and below 0°, positive and negative velocities of displacement along a line, and so on). Such magnitudes are expressed by corresponding positive and negative numbers. If a and b are positive numbers and $a < b$, then $-a > -b$, and any positive number, including zero, is greater than any negative number.

All rational and irrational numbers are arranged in a certain definite order, according to their magnitudes. All these numbers form the aggregate of *real numbers.*

We shall note one circumstance in connection with the representation of real numbers by decimal fractions. We can write an infinite decimal fraction with nine recurring in place of any given finite decimal fraction. For example: $3.16 = 3.1599\ldots$ If finite decimal fractions are not used, an accurate one-to-one correspondence is then obtained between real numbers and infinite decimal fractions, i. e. to every real number, except zero, there corresponds a definite infinite decimal fraction and to every infinite decimal fraction there corresponds a definite real number. Negative numbers can be associated with corresponding infinite decimal fractions with the minus sign in front.

The four primary operations can be carried out in the domain of real numbers, except division by zero. The root of odd degree of any given real number always has one specific value. The root of even degree of a positive number has two values, which differ only in sign. The root

of even degree of a negative real number has no meaning in the domain of real numbers. The rigorous theory of real numbers and the operations on them is given later in small type in [40].

The number expressing a given magnitude is called its arithmetic or absolute value when associated with the + sign. The absolute value of the magnitude expressed by the number a, or in other words, the absolute value of the number a, is denoted by the symbol $|a|$. Thus we have:

$$|a| = a, \text{ if } a \text{ is a positive number,}$$
$$|a| = -a, \text{ if } a \text{ is a negative number.}$$

It can easily be shown that the absolute value of the sum $|a+b|$ is equal to the sum of the absolute values of the parts, $|a| + |b|$, only if the parts have the same sign; otherwise, it will be less, so that we have:

$$|a+b| \leqslant |a| + |b|.$$

For example, the absolute value of the sum of the numbers $(+3)$ and (-7) is equal to four, but the sum of the absolute values of the parts is equal to ten.

Similarly, it can be shown that

$$|a-b| \geqslant |a| - |b|,$$

on the assumption that $|a| \geqslant |b|$.

The absolute value of the product of any number of factors is equal to the product of the absolute values of these factors, and the absolute value of a quotient is equal to the quotient of the absolute values of numerator and denominator, i.e.:

$$|abc| = |a|\cdot|b|\cdot|c| \text{ and } \left|\frac{a}{b}\right| = \frac{|a|}{|b|}.$$

3. Constants and variables. The magnitudes studied in mathematics are divided into two classes: *constants* and *variables.*

A magnitude is called a constant when it retains the same (invariable) value in a given investigation; a magnitude is a variable when, for one reason or another, it can assume different values in a given investigation.

It is evident from these definitions that the concepts of constant and variable are largely a matter of convention and depend on the circumstances in which the given phenomenon is studied. A magnitude

that may be considered as a constant under certain conditions, can become a variable under different conditions, and conversely.

For instance, it is important to know, when measuring the weights of bodies, whether the weighing operations are carried out at the same point of the earth's surface, or at different points: if the measurements are made at the same point, the acceleration due to gravity, on which the weight depends, will remain constant, and differences in weight between different bodies will depend only on their masses. On the other hand, if the measurements are made at different points on the earth's surface, the acceleration due to gravity cannot be looked on as a constant, since the centrifugal force due to the rotation of the earth must be considered. As a result of this factor, the same body will weigh less at the equator than at the poles, as can be observed if a spring-balance is used, instead of a lever-balance.

Similarly, the length of the rods used in the construction of some technical device can be looked on as invariable for the purposes of rough calculation. When greater accuracy is needed, so that regard is taken of the effect of temperature on the measurement, the length of a rod becomes variable, with the natural result that all calculations become more complex.

4. Interval. The character of the change of a variable can be extremely diverse. A variable can assume either all possible real values, without limits (time t, for example, calculated from some definite initial moment, can assume all possible values, both positive and negative), or its values are limited by certain *inequalities* (absolute temperature $T°$, for example, must be greater than $-273°$ C); or finally, a variable can assume only certain, and not all possible, types of value (only integral, in the case of the population figure for a given year or for the number of molecules in a given volume of gas, or only commensurable with a given unit and so on).

We shall note some of the most common ways in which variables change in theory and practice.

If the variable x can assume every real value permitted by the condition $a \leqslant x \leqslant b$, where a and b are fixed real numbers, we say that *x varies in the interval* (a, b). Such an interval, including its ends, is sometimes referred to as a *closed interval.* If the variable x can assume all values in the interval (a, b) except its ends, i.e. $a < x < b$, we say that *x varies inside the interval* (a, b). Such an interval, with its ends excluded, is referred to as an *open interval.* Furthermore, the

domain of variation of x can be an interval, closed at one side and open on the other: $a \leqslant x < b$ or $a < x \leqslant b$.

If the domain of variation of x is defined by $a \leqslant x$, we say that x varies in the interval $(a, +\infty)$, which is closed on the left and open on the right. Similarly, if $x \leqslant b$, we have the interval $(-\infty, b)$, open on the left and closed on the right. If x can assume any real value, we say that x varies in the interval $(-\infty, +\infty)$, open on both sides.

5. The concept of function. We are concerned in most applications not with one variable, but with several variables at once.

Let us take the example of a certain quantity of air, say 1 kg; the variables defining its state are: its pressure p (kg/m^2), the volume v (m^3) which it occupies; its temperature t°C. Let us assume for the moment that the temperature of the air is maintained at 0°C; the number t is then a constant, equal to zero. The only remaining variables are p and v. If the pressure p changes, then the volume v changes; for example, if the air is compressed, the volume decreases. We can change p arbitrarily (at least within the limits technically attainable), in which case we can refer to p as an *independent variable;* for every pressure p, there is evidently a completely defined volume. There must thus be a law which enables the corresponding volume v to be found for every value of p. This is, of course, Boyle's law, which says that the volume occupied by a gas at constant temperature is inversely proportional to the pressure.

Applying this law to our kilogram of air, the relationship between v and p can be put in the form of an *equation:*

$$v = \frac{273 \times 29.27}{p}.$$

The variable v is in this case called a *function* of the independent variable p.

Turning from this particular example, we can say that, theoretically speaking, *an independent variable is characterized by a large number of possible values, its value being any one chosen arbitrarily from all these possible values.* The independent variable x, for example, can have a set of values consisting of the interval (a, b), or the interior of this interval, i.e. the independent variable x can take any value satisfying the condition $a \leqslant x \leqslant b$, or $a < x < b$. It might be the case that x takes any integral value, etc. In the example quoted above, p had the role of independent variable, and the volume v was a function of p. We shall now define a function theoretically.

DEFINITION. *A quantity y is called a function of the independent variable x, if for any given value of x (from all its possible values) there corresponds a definite value of y.*

Thus, if y is a function of x, defined in the interval (a, b), this means that there is a corresponding definite value of y for any value of x from this interval.

It is mostly just a matter of convenience which of two magnitudes, x or y, is to be taken as the independent variable. In our example, we could have changed the volume v arbitrarily and defined the pressure p each time, thus having v as the independent variable and p as a function of v. We obtain an expression for p as a function of the independent variable by solving the equation given above in terms of p:

$$p = \frac{273 \times 29.27}{v}.$$

What has been said in regard to two variables is extended without difficulty to the case of any desired number of variables; and we can distinguish here between *independent variables, and dependent variables or functions.*

Returning to our example, let us assume that the temperature is no longer 0°C, but can change. Boyle's law must now be replaced by the more complex relationship of Clapeyron:

$$pv = 29.27\,(273 + T),$$

which shows that, when studying the gaseous state, only two of the magnitudes p, v and T can be changed arbitrarily, the third being fully defined when the values of these two are given. We can take p and T as independent variables, for example, in which case v is a function of them:

$$v = \frac{29.27\,(273 + T)}{p}.$$

Or similarly, v and T can be independent variables, and p a function of them.

Let us take another example. The area S of a triangle is given in terms of the lengths of sides a, b, c, by the formula:

$$S = \sqrt{p\,(p-a)\,(p-b)\,(p-c)},$$

where p is half the perimeter of the triangle:

$$p = \frac{a+b+c}{2}.$$

The sides a, b, c can vary arbitrarily, provided only that each side is greater than the difference, and less than the sum, of the other two. The variables a, b, c are thus *independent variables restricted by certain inequalities, whilst S is a function of them.*

We can also fix two sides, say a, b, and the area S of the triangle; using the formula:

$$S = \frac{1}{2} ab \sin C,$$

where C is the angle between sides a, b; we can then find C. The magnitudes a, b, S are the independent variables here, and C the function. The variables a, b, S must be restricted in this case by the condition:

$$\sin C = \frac{2S}{ab} \leqslant 1.$$

It may be noted that we obtain two values for C in this example, depending on whether we take the acute or obtuse angle, both of which have the same sine:

$$\sin C = \frac{2S}{ab}.$$

We meet here the concept of a *many-valued function*, about which more will be said below.

6. The analytic method of representing functional relationships. Every law of nature, connecting certain phenomena with others, establishes a *functional relationship* between magnitudes.

There are many ways of representing a functional relationship, but the most important are the three following: (1) the analytic, (2) the tabular, and (3) the graphical, or geometrical method.

We say that a functional relationship between magnitudes, or more simply, a *function*, is *represented analytically*, if these magnitudes are connected with each other by *equations*. These equations contain the magnitudes, subject to the various mathematical operations: addition, subtraction, division, taking logarithms etc. We always arrive at an analytic representation of a function on studying a problem *theoretically*: which means that, having established basic premises, we make use of mathematical analysis and obtain results in the form of mathematical formulae. For instance, in celestial mechanics, all the possible motions, positions and interactions of the heavenly bodies are deduced from a single basic law, that of universal gravitation.

If we have a direct expression for the function (i.e. the dependent variable) in terms of mathematical operations on the independent

variables, we say that the function is given explicitly. The expression for the volume v of a gas at constant temperature in terms of the pressure is an example of an explicit function (of one independent variable):

$$v = \frac{273 \times 29.27}{p}.$$

Similarly, the expression for the area S of a triangle in terms of the sides:

$$S = \sqrt{p(p-a)(p-b)(p-c)},$$

is an example of an explicit function of three independent variables. Another example may be given of an explicit function of one independent variable:

$$y = 2x^2 - 3x + 7. \tag{1}$$

It is often inconvenient or impossible to write down the formula expressing a function in terms of the independent variables. We can write briefly instead:

$$y = f(x).$$

This notation means that y is a function of the independent variable x, and f symbolizes the dependence of y on x. Of course another letter can be used in place of f. If we are considering several different functions of x, several different letters must be used to express symbolically the dependence on x:

$$f(x),\quad F(x),\quad \varphi(x),\ \text{etc.}$$

This notation is not only used when the function is given analytically, but is used in the general case of functional dependence, as defined in [5].

Use is made of an analogous abbreviated notation for functions of several independent variables:

$$v = F(x, y, z).$$

Here, v is a function of the variables x, y, z.

We obtain particular values of functions by giving the independent variables particular values and carrying out the operations indicated by the symbols $f, F, \ldots$ For example, the particular value of the function (1) for $x = 1/2$ is:

$$y = 2 \cdot \left(\frac{1}{2}\right)^2 - 3 \cdot \frac{1}{2} + 7 = 6.$$

In general, the particular value for $x = x_0$ of some function $f(x)$ is denoted by $f(x_0)$ and similarly for functions of several variables.

It is as well not to confuse the general concept of a function, which we gave in [5], with the concept of the analytic expression of y in terms of x. Reference is made in the general definition of a function only to some law, in accordance with which there is a corresponding definite value of y for any one of the set of possible values of the variable x. No analytic expression (formula) for y in terms of x is assumed here. It may be further remarked that a function can be defined by different analytic expressions for different portions of the domain of variation of the independent variable. For example, we can define the function y in the interval $(0, 3)$ in the following way: $y = x + 5$ for $0 \leqslant x < 2$, and $y = 11 - 2x$ for $2 \leqslant x \leqslant 3$. A corresponding value of y is in this case defined for any given x in the interval $(0, 3)$, which agrees with the definition of a function.

7. Implicit functions. A function is called implicit when we have no direct analytic expression for it in terms of the independent variables, but only an equation relating its values to those of the independent variables. For instance, if a variable y is related to a variable x by the equation:

$$y^3 - x^2 = 0,$$

y is an implicit function of the independent variable x; or on the other hand, x can be reckoned an implicit function of the independent variable y.

An implicit function v of several independent variables $x, y, z, \ldots$ is defined in general by an equation:

$$F(x, y, z, \ldots, v) = 0.$$

We can only compute the value of this function by solving the equation with respect to v, thus putting v in the form of an explicit function of $x, y, z, \ldots$:

$$v = \varphi(x, y, z, \ldots).$$

In the above example, y is expressed in terms of x as:

$$y = \sqrt[3]{x^2}.$$

However, it is by no means essential to solve the equation to obtain the various properties of the function v; more often than not, an

implicit function can be studied quite well from the equation that defines it, without attempting a solution.

The volume v of a gas, for example, is an implicit function of the pressure p and temperature T, defined by the equation:

$$pv = R(273 + T).$$

The angle C between sides a and b of a triangle of area S is an implicit function of a, b, S, defined by the equation:

$$ab \sin C = 2S.$$

8. The tabular method. The analytic method of representing a function is primarily used in theoretical work, when it is a matter of general laws. In practical work, which requires the computing of a large number of particular values of different functions, an analytic presentation is often unwieldy, since it means carrying out all the necessary calculations in each case.

To avoid this, particular values of the most commonly occurring functions are computed for a large number of particular values of the independent variables, and *tables* are compiled.

For example, there are tables of values of the functions:

$$y = x^2;\ \ \frac{1}{x};\ \ \sqrt{x};\ \ \pi x;\ \frac{1}{4}\pi x^2; \log_{10} x;\ \ \log_{10} \sin x;\ \ \log_{10} \cos x;\ \ \text{etc.}$$

Such functions are constantly met with in practical work. Very useful tables have also been compiled for more complicated functions, such as Bessel functions, elliptic functions etc. There are further tables for functions of more than one variable, the simplest example being the ordinary multiplication table, i.e. a table of values of the function $z = xy$ for different integral values of x and y.

Occasionally, the tables only give the function for particular values of the independent variables, adjacent to those values for which the function is required; in order to make it possible to use the tables in this case, various rules of *interpolation* exist; one such rule is given in school courses on using logarithmic tables (proportional parts).

Tables have special importance when they represent a function, the analytic expression for which is *unknown*; this is the case when an *experiment* is carried out. Every experimental investigation aims at discovering hidden functional relationships, and experimental results are set in the form of a *table*, relating the different values of the magnitudes investigated in the experiment.

9. The graphical method of representing numbers. Passing to the graphical method of representing functional relationships, we shall begin with the representation of a single variable.

Every number x can be represented by a certain interval. Having settled once for all on the choice of the unit of length, it is sufficient to construct an interval, the length of which is exactly equal to the given number x. Thus, every magnitude can be represented geometrically by an *interval*, as well as expressed by a *number*.

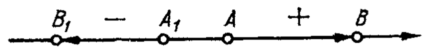

FIG. 1

So as to be able to represent negative numbers, we agree to cut off intervals along the same straight line, thus attributing a definite direction to them (Fig. 1). We agree further to denote every interval by a symbol $\overline{AB}$, the point A being called the origin, and the point B the end of the interval.

If the direction from A to B coincides with the direction of the line, the interval represents a positive number; whilst if the direction from A to B is opposite to that of the line, it represents a negative number ($A_1 B_1$ in Fig. 1). The *absolute* value of the number concerned is expressed by the length of the interval representing it, irrespective of direction.

The length of the interval $\overline{AB}$ will be denoted by $|\overline{AB}|$; if $\overline{AB}$ represents a number x, we shall simply write:

$$x = \overline{AB};\ |x| = |\overline{AB}|.$$

We can make things more definite by agreeing once and for all to locate the origin of all intervals at a previously chosen point O of

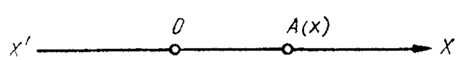

FIG. 2

the line. Every interval $\overline{OA}$, and hence the number x that it represents, will now be fully defined by the *point* A, its end (Fig. 2). Conversely, given the number x, we can define the magnitude and direction of the interval $\overline{OA}$, and hence its end A.

Thus, *if we take a straight line $X'X$ (axis) in a given direction, and mark out on it a point O (origin), a corresponding point A of this line will be defined for every real number x, the interval $\overline{OA}$ being measured by the number x. Conversely, for every point A of the axis there is a corresponding fully defined real number x, measuring $\overline{OA}$. This number x is called the abscissa of the point A; if we want to denote that the point A has abscissa x, we shall write $A(x)$.*

If the number x varies, the point A representing it moves on the axis. The concept of interval established above becomes easy to visualize with this *graphical representation of the number x*. If x varies in the interval $a \leqslant x \leqslant b$, the corresponding point on the axis $X' X$ can be situated anywhere in the interval, the ends of which have abscissae a and b.

If we were limited to rational numbers alone, there would be no abscissa corresponding to the point A when the interval $\overline{OA}$ proved incommensurable with the chosen unit, in other words, rational numbers alone do not fill in all the points of a straight line. This filling is achieved by introducing irrational numbers. The proposition mentioned above is basic to the graphical representation of a single variable: a corresponding real number is defined for every point of the axis $X' X$, and conversely, a corresponding point of the axis $X' X$ is defined for every real number.

Let us take two points on the axis $X' X$: the point A_1 with abscissa x_1, and point A_2 with abscissa x_2. The number x_1 will correspond to the interval $\overline{OA_1}$, and number x_2 to $\overline{OA_2}$. Taking all the possible mutual arrangements of points A_1 and A_2, it is easy to see that the interval $\overline{A_1A_2}$ corresponds to the number $(x_2 - x_1)$, so that the length of the interval is equal to the absolute value of the difference $(x_2 - x_1)$:

$$|\overline{A_1 A_2}| = |x_2 - x_1|.$$

If, for example, $x_1 = -3$ and $x_2 = 7$, the point A_1 lies to the left of O at a distance equal to 3, and point A_2 lies to the right of O at a distance equal to 7. The section $\overline{A_1A_2}$ will have length 10 and will be directed along the axis $X' X$, i.e. corresponding to it we have the number $10 = 7 - (-3) = x_2 - x_1$. We leave it to the reader to analyse other possible arrangements of the points A_1 and A_2.

10. Coordinates. We saw above, that the position of a point on a line $X' X$ can be defined by a real number x. We now indicate the analogous method of defining the position of a point on a plane.

We shall take two mutually perpendicular axes $X'X$ and $Y'Y$ on a plane, with their point of intersection O as origin on each (Fig. 3). The positive directions on the axes are shown by arrows. We have a real number, denoted by the letter x, corresponding to a point of the axis $X'X$. Similarly, a real number denoted by y corresponds to a point of the axis $Y'Y$. If specific values are assigned to x and y, points A and B are defined on axes $X'X$ and $Y'Y$; knowing A and B, we can construct the point M as the intersection of lines parallel to the axes and passing through A and B.

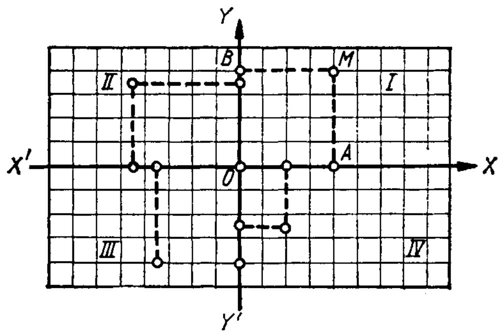

FIG. 3

To each pair of values x, y, there corresponds a single fully defined position of the point M on the plane of the figure.

Conversely, to each point M of the plane there corresponds a fully defined pair of values of x, y, corresponding to the points at which lines through the point M parallel to the axes intersect the axes $X'X$ and $Y'Y'$.

With the directions of the axes shown in Fig. 3, x is to be reckoned positive or negative, depending on whether A lies to the right or left of point O; similarly, y is positive or negative, depending on whether B lies above or below, point O.

The magnitudes x, y defining the position of point M in the plane, and defined in turn by point M, are called the coordinates of M. The axes $X'X$, $Y'Y$ are called the coordinate axes, the plane of the figure is the plane of coordinates XOY, and the point O is the origin of coordinates.

Magnitude x is called the abscissa, and y the ordinate, of the point M. We shall specify the point M by its coordinates by writing: $M(x, y)$.

This method of representation is called the method of rectangular coordinates.

The signs of the coordinates of the point M when situated in different quadrants of the axes (I–IV) (Fig. 3) can be shown in a table:

M	I	II	III	IV
x	+	−	−	+
y	+	+	−	−

It is obvious that the coordinates x, y of M are equal to the *distances* of M from the coordinate axes, associated with the corresponding signs.

11. Graphs. The equation of a curve. We return to x and y, representing the point M. Let x and y be connected by a functional relationship This means that, on varying x (or y) arbitrarily, a corresponding value of y (or x) can be found each time. Every such pair of values x and y corresponds to a definite position of the point M on the plane XOY; when the values vary, the point M moves over the plane and thus traces out a certain curve (Fig. 4), which is called a *graphical representation* (or simply a *graph* or *diagram*) *of the functional relationship concerned.*

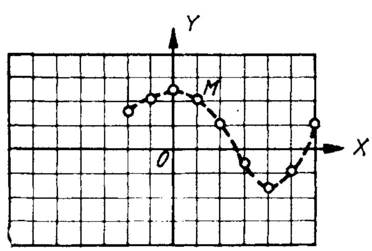

FIG. 4

If the relationship is given *analytically* as an *equation* in explicit form:

$$y = f(x),$$

or in implicit form:

$$F(x, y) = 0,$$

we call this the *equation of the curve*, whilst the curve is the *graph of the equation*. A curve and its equation are simply different expressions of the same functional relationship, i.e. *all points, the coordinates of which satisfy the equation of a curve, lie on this curve, and conversely, the coordinates of all points lying on the curve satisfy its equation.*

If the equation of a curve is given, the curve itself can be constructed more or less accurately on a sheet of graph paper (more strictly, any desired number of points lying on the curve can be constructed); the more points are taken, the more evident becomes the shape of the curve. This method is called *plotting a curve.*

The choice of scale is important in plotting curves. Different scales can be chosen for x and y. The plane is taken as a sheet of paper, ruled into *squares* or *rectangles*, depending on whether the scales of x and y are the same or different. It is assumed below that the scales of x and y are the same.

The reader is recommended at this point to plot some curves of simple functions, and to vary the scales of x and y.

The concepts introduced above of the coordinates of a point M, of the equation of a curve and the graph of an equation, establish an intimate connection between algebra and geometry. On the one hand, we can represent and study an analytic relationship by a visual, geometrical method, on the other hand; it becomes possible to solve geometrical problems with purely algebraic operations, including here the fundamental work of *analytic geometry*, first undertaken by Descartes.

In view of its extreme importance, we shall formulate again the facts that lie at the basis of analytic geometry. If we mark out two coordinate axes in a plane, *every point of the plane corresponds to a pair of real numbers, the abscissa and ordinate of the point, and conversely, every pair of numbers corresponds to a definite point of the plane, the first number being its abscissa and the second its ordinate. A curve on the plane corresponds to a functional relationship between x and y, that is, to an equation containing x and y which is satisfied if, and only if, x and y can be replaced by the coordinates of some point of the curve. Conversely, an equation containing two variables x and y has a corresponding curve, made up of the points of the plane whose coordinates, when substituted for x and y, satisfy the equation.*

We now turn to the study of the graphs of simple functions. We again note that if we have a functional relationship given by an equation in explicit or implicit form:

$$y = f(x), \text{ or } F(x, y) = 0,$$

then the curve in the plane of axes $X'X$, $Y'Y$ corresponding to this equation is called the *graph of the equation* or the *graph of the function* defined by the equation. The abscissae and ordinates of points of this curve give mutually corresponding values of the variables x and y, connected by the functional relationship.

Graphs are drawn automatically in recording devices; the variable x is usually time; y is the magnitude, the variation of which with time interests us: for example, barometric pressure (barograph) or temperature

(thermograph). The indicator which records the relationship between the volume and pressure of the gas enclosed in a steam or gas engine is worth mentioning.

12. Linear functions. The simplest function, which at the same time has extremely important applications, is the *polynomial of the first degree:*

$$y = ax + b, \tag{2}$$

where a and b are given constant numbers. We shall see that the graph of this function is a straight line. It is called a *linear function.* We shall first consider the case of b equal to zero. The function then has the form:

$$y = ax. \tag{3}$$

This expresses the fact that the variable y is directly proportional to the variable x, the constant coefficient a being called the coefficient of proportionality.

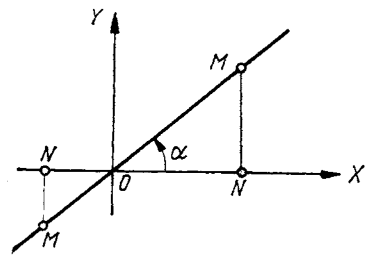

FIG. 5

Turning to the figure (Fig. 5), we see that equation (3) expresses the following geometric property of the graph in question: whatever point M we take on it, the ratio of the ordinate $y = \overline{NM}$ of this point to its abscissa $x = \overline{ON}$ is equal to the constant a. Since, on the other hand, this ratio is equal to the tangent of the angle α between the segment $\overline{OM}$ and the axis $\overline{OX}$, it is evident that the geometrical locus of M is *a straight line, passing through the origin of coordinates* O at an angle α (or $\pi + \alpha$) to the axis $\overline{OX}$.

Angle α is reckoned counter-clockwise from the direction OX.

The geometrical importance of the coefficient a in equation (3) is simultaneously revealed: a is the tangent of the angle α between the axis OX and the straight line corresponding to this equation, a being therefore called the *slope* of the straight line. It may be noted that if a is a negative number, the angle α is obtuse, and the corresponding line is as shown in Fig. 6.

Let us now return to the general case of a linear function, viz, to equation (2). The ordinate y of the graph of this equation differs from the corresponding ordinate of the graph of equation (3) by the constant amount b. Thus, we immediately obtain the graph of equation (2), if the graph of equation (3) shown in Fig. 5 (for $a > 0$) is displaced

parallel to the axis OY through a distance b: upwards for b positive, downwards for b negative. We obtain a straight line, parallel to the initial line, and cutting off a segment $\overline{OM_0} = b$ on the axis OY (Fig. 7).

Thus, *the graph of function (2) is a straight line, the coefficient a being equal to the tangent of the angle that the line makes with the axis OX, and the constant term b equal to the segment cut off by the line on the axis OY, measured from the origin O.*

Conversely, given any straight line L, not parallel to the axis OY, its equation can easily be written in the form (2). In accordance with the above, it is sufficient to take the coefficient a equal to the

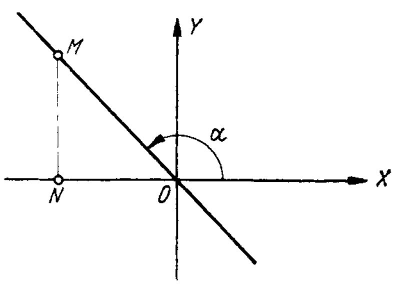

Fig. 6

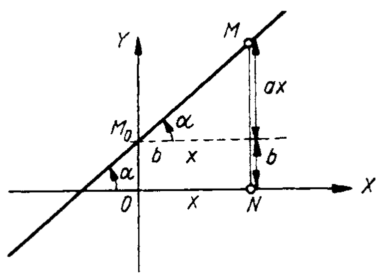

Fig. 7

tangent of the angle of inclination of this line to the axis OX, and b equal to the segment that it cuts off on the axis OY.

We shall note a particular case, which presents a well-known peculiarity. Let $a = 0$. Equation (2) gives for all x:

$$y = b, \tag{2_1}$$

i.e. a "function" of x is obtained, such that its value b remains the same for all values of x. It is obvious that the graph of equation (2_1) is a straight line, parallel to the axis OX and displaced from it by a distance $|b|$ (upwards for $b > 0$, and downwards for $b < 0$). To avoid special reservations, we shall sometimes say that equation (2_1) also defines a function of x.

13. Increment. The basic property of a linear function. We shall establish one important new concept, with which we shall often be concerned in the study of functional relationships.

The difference between the final and initial values of the independent variable x on transition from the initial value x_1 to the final value x_2

is called the increment of x, equal to $x_2 - x_1$. The difference between the final and initial values of the function $y = f(x)$ is called the corresponding increment of the function:

$$y_2 - y_1 = f(x_2) - f(x_1). \tag{3}$$

These increments are often denoted by:

$$\Delta x = x_2 - x_1,\ \Delta y = y_2 - y_1.$$

It may be noted here that the increment can be either positive or negative, and that the magnitude receiving the "increment" does not necessarily increase.

It must be pointed out that the symbol Δx has to be regarded as a single entity in denoting the increment of x.

We shall consider the case of a linear function, when

$$y_2 = ax_2 + b \text{ and } y_1 = ax_1 + b.$$

Subtracting term by term, we have:

$$y_2 - y_1 = a(x_2 - x_1) \tag{4}$$

or

$$\Delta y = a\Delta x.$$

This equality shows that *the linear function* $y = ax + b$ has the property that the increment of the function $(y_2 - y_1)$ is proportional to the increment of the independent variable $(x_2 - x_1)$, the coefficient of proportionality being equal to a, i.e. *to the slope of the graph of the function.*

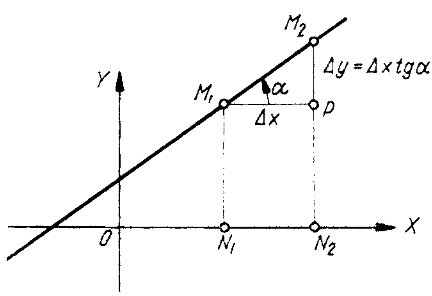

Fig. 8

Turning to the graph itself (Fig. 8), corresponding to the increment of the independent variable we have the segment $\overline{M_1P} = \Delta x = x_2 - x_1$, and corresponding to the increment of the function, $\overline{PM_2} = \Delta y = y_2 - y_1$; and formula (4) follows at once from considering the triangle M_1PM_2.

We shall now assume that a certain function has the above property of proportionality of the increments of the independent variable and

of the function, expressed by (4). It follows from this formula that:

$$y_2 = a(x_2 - x_1) + y_1$$

or

$$y_2 = ax_2 + (y_1 - ax_1).$$

We shall take the initial values of the variables x_1 and y_1 as completely defined, and denote the difference $(y_1 - ax_1)$ by the single letter b:

$$y_2 = ax_2 + b.$$

Since we can take any final values of the variables x_2 and y_2, we can simply write in place of the letters x_2 and y_2 the letters x and y, and rewrite the above equality in the form:

$$y = ax + b,$$

i.e. *every function having the above property of proportionality of the increments is a linear function $y = ax + b$, where a is the coefficient of proportionality.*

A linear function and its graph, a straight line, can thus serve to represent any natural law in which the increments of the magnitudes concerned are proportional, as is very often the case.

14. Graph of uniform motion. This is one of the most important applications, which gives a *mechanical* interpretation of the equation of a straight line and its coefficients. If the point P moves along a certain path (trajec-

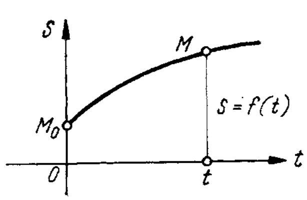

FIG. 9

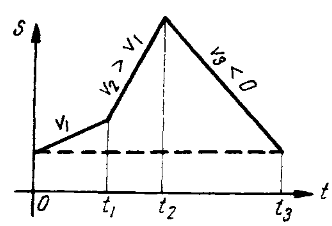

FIG. 10

tory), its position is fully defined by the distance, measured from either side along the trajectory from a given point A of it to the point P. This distance, i.e. the arc AP, is called the path traversed, and is denoted by the letter s; s can be both positive and negative, its values on one side of the initial point A being reckoned positive, and on the other side, negative.

The path s traversed is a certain function of time t; taking time as the independent variable, we can draw a *graph of the motion*, i.e. a graph of the functional relationship (Fig. 9):

$$s = f(t);$$

this is not to be confused with the trajectory itself.

The motion is called *uniform* if the path traversed by a point in any given interval of time is proportional to this interval, in other words, if the ratio

$$\frac{s_2 - s_1}{t_2 - t_1} = \frac{\Delta s}{\Delta t}$$

of the path traversed in the interval from t_1 to t_2, to the size of this interval, is a constant; this constant is called the *velocity* of the motion and is denoted by v.

It is clear from the above that the *equation of the graph* of uniform motion has the form:

$$s = vt + s_0;$$

the graph itself is a straight line, the slope of which is equal to the velocity whilst the initial ordinate s_0 is the initial value of the path s traversed, i.e. *the value of s* at $t = 0$.

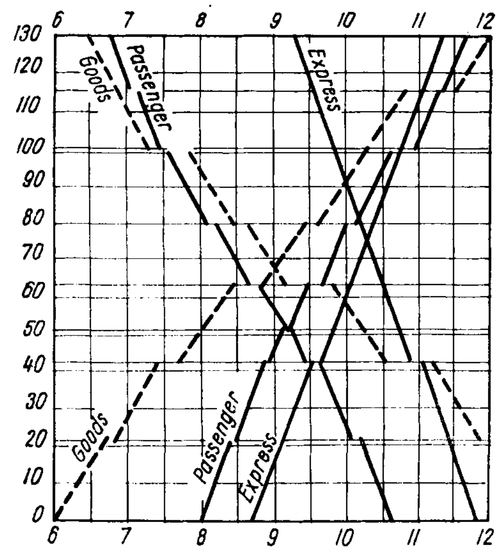

FIG. 11

Figure 10 shows the graph of the motion of a point P, moving with constant velocity v_1 in a positive direction from the instant 0 to the instant t_1 (acute angle with axis of t), then with higher constant velocity v_2 in the same direction (larger acute angle) to the instant t_2, then with negative constant velocity v_3 (in the opposite direction, obtuse angle) back to its initial position. In the case of several points, all moving in the same trajectory (for instance, when making up schedules of the movement of trains or trams), this *graphical method* is particularly suitable as a practical means of determining the encounters of the moving points, and in general for reviewing the movements as a whole (Fig. 11).

15. Empirical formulae. The simplicity of constructing a straight line and of thus expressing a law of proportionality between the increment of a function and that of the independent variable makes a straight line graph an extremely convenient means of arriving at an empirical law, i.e. one following directly from the experimental data, without special theoretical investigation.

We represent graphically on a sheet of millimetre paper the table of experimental results and thus obtain a series of points. If we wish to obtain an approximate empirical formula so as to study a functional relationship in the form of a linear function, we now draw a straight line, which, if it does not pass through all the points given (which, of course, is almost always an impossibility), is made at the least to pass between these points in such a way that as nearly the same number of points as possible appear on one side of the line as on the other, all the points being reasonably close to it. Error theory, and observation theory, study more accurate ways of drawing the line mentioned, as also for judging the degree of error arising with such

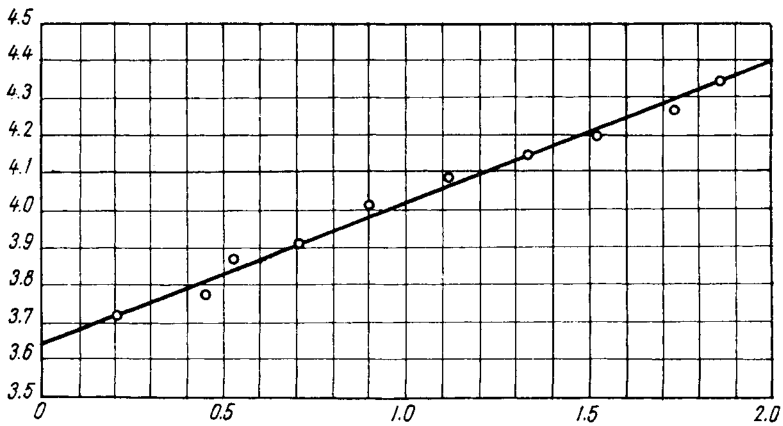

FIG. 12

an approximation. In less accurate investigations, however, such as in technology, the drawing of the empirical line is most simply carried out by the "taut string" method, the nature of the method being evident from the name. Having obtained the line, we use direct measurement to determine its equation:

$$y = ax + b\,,$$

which gives the required empirical formula. When obtaining this formula we must not lose sight of the fact that different scales are very often in use for the magnitudes x and y, i.e. *lines with the same slope on the axes OX, OY, may represent different numbers.* In this case, the slope a will not be equal to the tangent of the angle between the line and the axis OX, but will differ from this by a factor numerically equal to the ratio of the units of length used in representing magnitudes x and y.

Example (Fig. 12).

x	0.212	0.451	0.530	0.708	0.901	1.120	1.341	1.520	1.738	1.871
y	3.721	3.779	3.870	3.910	4.099	4.089	4.150	4.201	4.269	4.350

The result is:

$$y \sim 0.375x + 3.65.$$

(Here, and subsequently, the sign $\sim$ denotes approximately equal to.)

16. Parabola of the second degree. The linear function

$$y = ax + b$$

is a particular case of an *integral function of the n-th degree* or a *polynomial of the n-th degree*:

$$y = a_0 x^n + a_1 x^{n-1} + \ldots + a_{n-1} x + a_n,$$

the simplest case of which, after the linear function, is the *polynomial of the second degree* ($n = 2$):

$$y = ax^2 + bx + c;$$

the graph of this function is called a *parabola of the second degree* or simply a *parabola.*

For the present, we shall consider only the simplest case of a parabola:

$$y = ax^2. \tag{5}$$

This curve can easily be plotted. Figure 13 shows the curves $y = x^2$ ($a = 1$) and $y = -x^2$ ($a = -1$). The curve corresponding to equation (5) is situated wholly above the axis OX for $a > 0$, and wholly below the axis OX for $a < 0$. The ordinate of this curve increases in absolute value when x increases in absolute value, the increase being the faster, the greater the absolute value of a. Figure 14 shows a series of graphs of the function (5) for different values of a, these values being marked in the figure against the corresponding parabolas.

Equation (5) contains only x^2, and hence does not vary on changing x to $-x$, i.e. if a given point (x, y) lies on the parabola (5), the point $(-x, y)$ also lies on it. The two points (x, y) and $(-x, y)$ are evidently symmetrical relative to the axis OY, i.e. one of them is the mirror image of the other relative to this axis. Thus, if the right-hand portion of the plane is turned through 180° about the axis OY and combined with the left-hand portion, the part of the parabola lying to the right of the axis OY will coincide with the part of the parabola lying to the left of this axis. In other words, axis OY is an *axis of symmetry of the parabola* (5).

The origin of coordinates is the lowest point of the curve for $a > 0$, and the highest point for $a < 0$, and is called the *vertex of the parabola.*

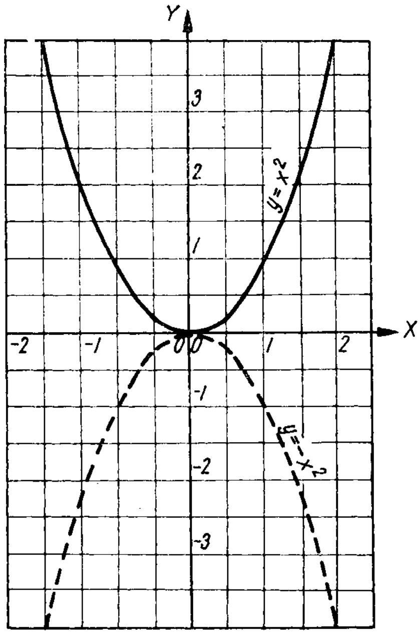

FIG. 13

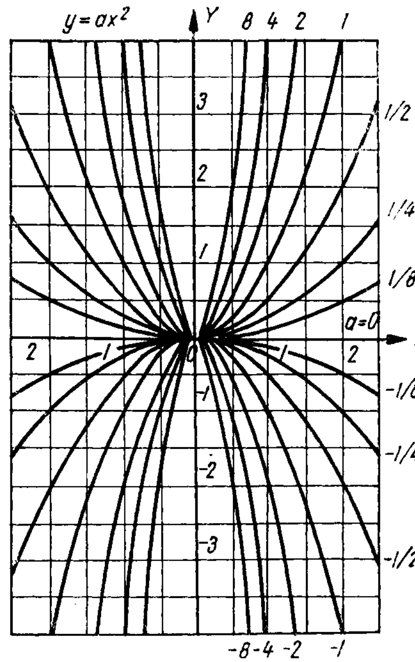

FIG. 14

The coefficient a is fully defined if one point $M_0\,(x_0, y_0)$ of the parabola is given, not at the vertex, since we have then:

$$y_0 = ax_0^2, \quad a = \frac{y_0}{x_0^2},$$

so that the equation of the parabola (5) now has the form:

$$y = y_0\left[\frac{x}{x_0}\right]^2. \tag{6}$$

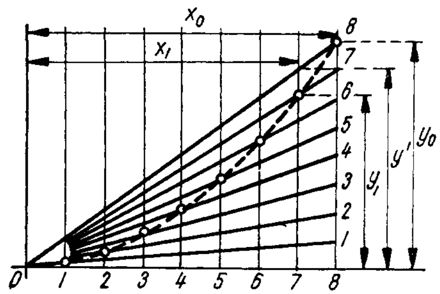

FIG. 15

There is an extremely simple method of constructing any desired number n of points of the parabola, given the vertex, the axis of symmetry and any point M_0 of it, not at the vertex.

We divide the abscissa and ordinate of the given point M_0 into n equal parts (Fig. 15) and draw a pencil of lines from the origin to the points of division of the ordinate. The points of intersection of these

lines with lines through the points of division of the abscissa and parallel to the axis OY are points of the parabola. In fact, we have by construction (Fig. 15):

$$x_1 = x_0 \cdot \frac{n-1}{n}, \qquad y' = y_0 \cdot \frac{n-1}{n},$$

$$y_1 = y' \cdot \frac{n-1}{n} = y_0 \left[\frac{n-1}{n}\right]^2 = y_0 \left[\frac{x_1}{x_0}\right]^2,$$

i.e. from (6), the point M_1 (x_1, y_1) also lies on the parabola. The proof is similar for the remaining points.

If we have two functions:

$$y = f_1(x) \text{ and } y = f_2(x)$$

and their corresponding graphs, the coordinates of the points of intersection of the graphs satisfy both the equations, i.e. the abscissae of the points of intersection are solutions of the equation:

$$f_1(x) = f_2(x).$$

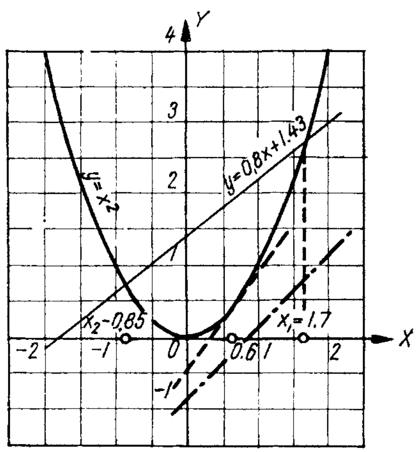

FIG. 16

This fact can readily be used solve a qudaratic equation approximately. Having constructed as accurately as possible the graph of the parabola

$$y = x^2 \tag{6_1}$$

on a piece of millimetre graph paper we can now find the roots of the quadratic equation

$$x^2 = px + q \tag{7}$$

as the abscissae of the points of intersection of the parabola (6_1) and the straight line $y = px + q$. Three cases are shown in Fig. 16, of two points of intersection, of one point (tangent to the parabola), and of no point of intersection.

17. Parabola of the third degree. The third degree polynomial:

$$y = ax^3 + bx^2 + cx + d,$$

has a graph in the form of a curve called a *parabola of the third degree.* We shall consider the simplest case of this curve:

$$y = ax^3. \tag{8}$$

For a positive, the signs of x and y are the same; for a negative, they are different. In the first case, the curve lies in the first and third quadrants of the coordinate axes, and in the second case, it lies

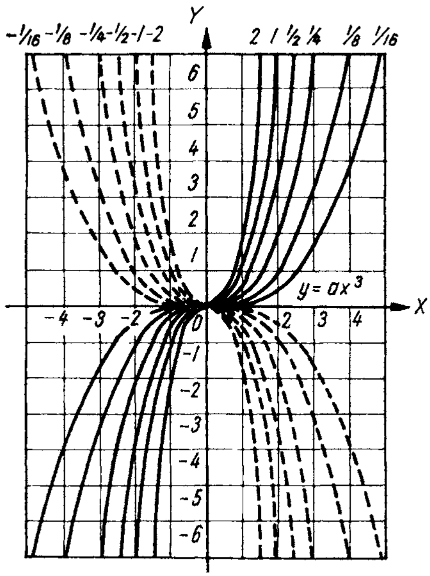

FIG. 17

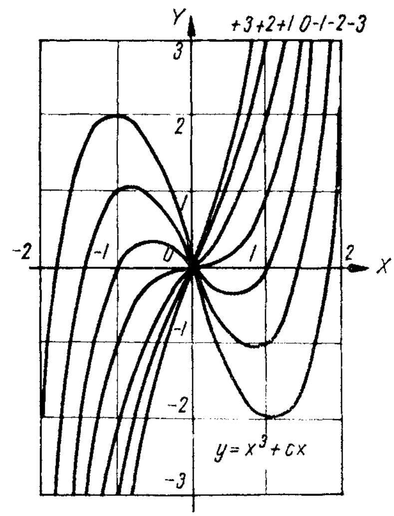

FIG. 18

in the second and fourth quadrants. Figure 17 illustrates the form of the curve for various values of a.

If x and y change simultaneously to $(-x)$ and $(-y)$, both sides of equation (8) change sign, and the equation is not essentially altered, i.e. if the point (x, y) lies on the curve (8), the point $(-x, -y)$ also lies on this curve. The points (x, y) and $(-x, -y)$ are evidently symmetrical relative to the origin O, i.e. the line joining them is bisected at O. It follows from this that every chord of the curve (8) that passes through the origin of coordinates O is bisected there. In other words: *the origin of coordinates O is the centre of the curve (8).*

A further particular case of a third degree parabola will be mentioned:

$$y = ax^3 + cx. \tag{9}$$

The right-hand side of this equation is the sum of two terms, and consequently, to construct the curve, it is sufficient to draw the straight line

$$y = cx \tag{10}$$

and take the sums of corresponding ordinates of graphs (8) and (10) directly from the figure. Figure 18 illustrates the various forms that the curve (9) assumes (with $a = 1$, and various c).

If the curve

$$y = x^3$$

is drawn, we obtain a convenient, though not too accurate, *method for solving graphically an equation of the third degree:*

$$x^3 = px + q \, ,$$

the roots of this equation being, in fact, the abscissae of the points of intersection of the curve $y = x^3$ with the line

$$y = px + q \, .$$

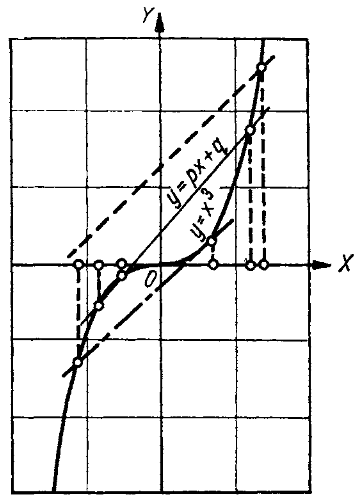

FIG. 19

As can be seen from the figure (Fig. 19), there may be one, two, or three points of intersection, but there must be at least one, i.e. *an equation of the third degree must have at least one real root.* A rigorous proof of this is given later.

18. The law of inverse proportionality. The functional relationship:

$$y = \frac{m}{x} \tag{11}$$

expresses the law of inverse proportionality between variables x and y. Variable y decreases by as many times as x increases. For $m > 0$, x and y have the same sign, i.e. the graph is located in the first and third quadrants of the coordinate axes; similarly, the graph is in the second and fourth quadrants, for $m < 0$. As x approaches zero, the absolute value of the fraction m/x becomes very large. Conversely, for large absolute values of x, the ratio m/x becomes small in absolute value.

Plotting this curve directly gives us Fig. 20, which shows curves (11) for various m, the full-line curves corresponding to $m > 0$, and the broken curves to $m < 0$, the corresponding value of m being noted against each curve. We see that each of the curves drawn, called *rectangular hyperbolas*, has *infinite branches*, the points on a given branch approaching the coordinate axis OX or OY with indefinite increase of the abscissa x or the ordinate y, respectively. These lines are called *asymptotes* to the hyperbola.

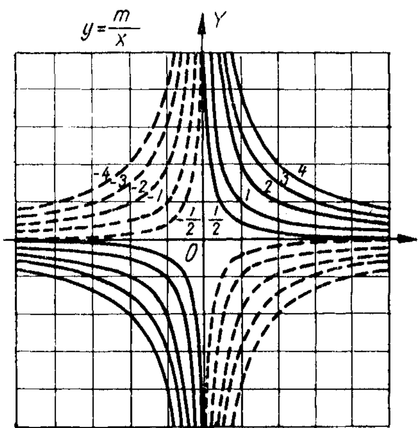

FIG. 20

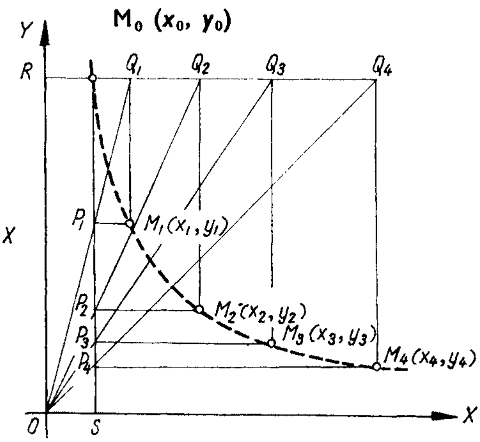

FIG. 21

The coefficient m in equation (11) is completely defined if any point $M_0(x_0, y_0)$ of the curve in question is given, since now:

$$x_0 y_0 = m;$$

equation (11) can now be written:

$$xy = x_0 y_0 \tag{12}$$

or

$$\frac{y}{x_0} = \frac{y_0}{x}.$$

Hence follows a *graphical method* for obtaining any required number of points of a rectangular hyperbola, given its asymptotes and any one point of it $M_0(x_0, y_0)$. Taking the asymptotes as coordinate axes, we draw an arbitrary pencil of lines OP_1, OP_2, ... from the origin, and mark off the points of intersection of these lines with the lines $y = y_0$ and $x = x_0$.

We draw new lines parallel to the axes through each pair of points of intersection lying on a given line of the pencil ; the points of inter-

section of these new lines are then points of the rectangular hyperbola (Fig. 21). This follows from the similarity of triangles ORQ_1 and OSP_1:

$$\frac{\overline{SP_1}}{\overline{OS}} = \frac{\overline{OR}}{\overline{RQ_1}} \quad \text{or} \quad \frac{y_1}{x_0} = \frac{y_0}{x_1},$$

i.e. the point $M_1(x_1, y_1)$ lies on the curve (12).

19. Power functions. The functions $y = ax$, $y = ax^2$, $y = ax^3$ and $y = m/x$ which we have studied above, are particular cases of a function of the form:

$$y = ax^n, \tag{13}$$

where a and n are any given constants. Function (13) is called in general a *power function*. We shall confine ourselves to drawing the curves for x positive and $a = 1$. Figures 22 and 23 show the graphs corresponding to various values of n.

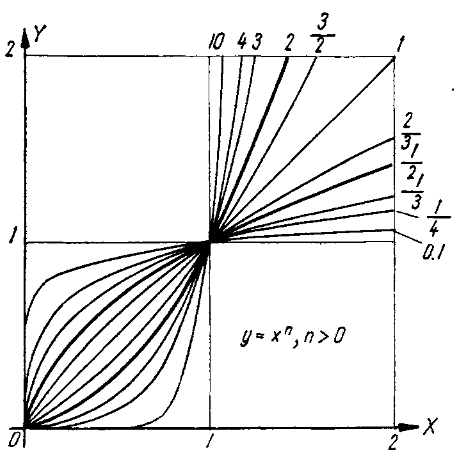

FIG. 22

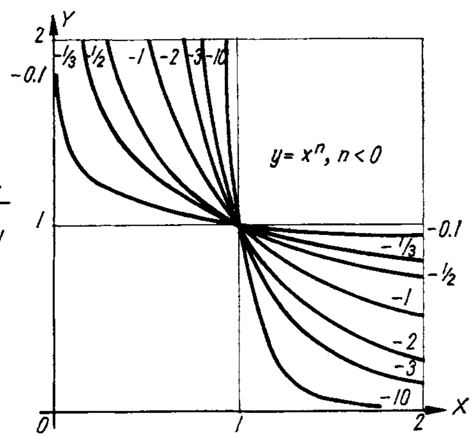

FIG. 23

The equation $y = x^n$ gives $y = 1$ at $x = 1$ for all values of n, i.e. all the curves pass through the point (1, 1). The curves rise for n positive and $x > 1$, the rate of rise increasing with the size of n (Fig. 22). The function $y = x^n$ is equivalent to a fraction for negative n (Fig. 23). Instead of $y = x^{-2}$, for instance, we can write $y = 1/x^2$. In these cases, the ordinate y diminishes with increase of x. The curves corresponding to equation (13) are sometimes referred to as *polytropic*. They are often encountered in thermodynamics.

It may be noted here, that we take the value of the radical as positive, for fractional n; for example, we take as positive $x^{\frac{1}{2}} = \sqrt{x}$.

The two constants a and n appearing in equation (13) are defined, provided two points of the curve $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$ are given, in which case we have:

$$y_1 = ax_1^n, \quad y_2 = ax_2^n; \tag{14}$$

we eliminate a by dividing one equation by the other:

$$\frac{y_1}{y_2} = \left[\frac{x_1}{x_2}\right]^n.$$

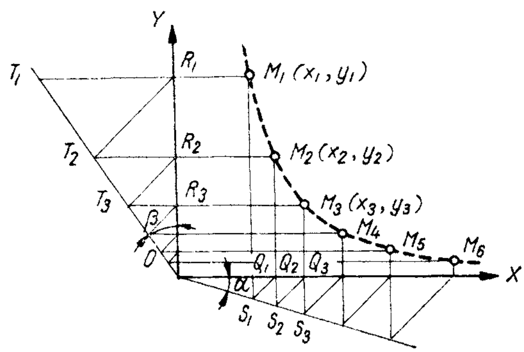

FIG. 24

Taking logarithms, we obtain n as

$$n = \frac{\log y_1 - \log y_2}{\log x_1 - \log x_2};$$

having found n, we obtain a from either of equations (14).

Figure 24 illustrates a *graphical method* of obtaining any required number of points of the curve (13), given two of its points $M_1(x_1, y_1)$ and $M_2(x_2, y_2)$. We draw two arbitrary lines through the point O at angles α and β to axes OX and OY respectively; we then take perpendiculars to the axes from the given points M_1, M_2, intersecting the arbitrary lines in points S_1, S_2; T_1, T_2, and intersecting the axes in points Q_1, Q_2; R_1, R_2. We now draw $R_2 T_3$ through R_2 parallel to $R_1 T_2$, and $S_2 Q_3$ through S_2 parallel to $S_1 Q_2$. Having finally drawn lines parallel to the corresponding axes through T_3 and Q_3, the intersection of these gives us the point $M_3(x_3, y_3)$ of the curve. Taking similar triangles, we have:

$$\frac{\overline{OQ_3}}{\overline{OQ_2}} = \frac{\overline{OS_2}}{\overline{OS_1}}; \quad \frac{\overline{OS_2}}{\overline{OS_1}} = \frac{\overline{OQ_2}}{\overline{OQ_1}};$$

i.e.

$$\frac{\overline{OQ_3}}{\overline{OQ_2}} = \frac{\overline{OQ_2}}{\overline{OQ_1}} \quad \text{or} \quad \frac{x_3}{x_2} = \frac{x_2}{x_1},$$

whence:

$$x_3 = \frac{x_2^2}{x_1}.$$

It can be shown in exactly the same way that

$$y_3 = \frac{y_2^2}{y_1}.$$

Using (14), we find:

$$y_3 = \frac{(ax_2^n)^2}{ax_1^n} = a\left(\frac{x_2^2}{x_1}\right)^n = ax_3^n,$$

i.e. the point (x_3, y_3) lies on the curve (13), as was required to be shown.

20. Inverse functions. We shall introduce a new concept, that of inverse function, in order to study further elementary functions. As was mentioned in [5], we are free to choose either x or y as independent variable in a functional relationship between them, the actual decision being made purely as a matter of convenience. Let us take a certain function $y = f(x)$, with x as independent variable.

The function defined by the same functional relationship $y = f(x)$, but having y as independent variable and x as the function

$$x = \varphi(y),$$

is called the inverse of the given function $f(x)$, this latter being often called the direct function.

The notation for the variables is not important in itself, so that, denoting the independent variable in both cases by the letter x, we can say that $\varphi(x)$ is the inverse of $f(x)$. For example, if the direct functions are

$$y = ax + b, \quad y = x^n,$$

the inverse functions are

$$y = \frac{x-b}{a}, \quad y = \sqrt[n]{x}.$$

Finding the inverse function from the equation of the direct function is called *inversion.*

Let us take the graph of the direct function $y = f(x)$. It can easily be seen that the same graph can serve as a graph of the inverse function $x = \varphi(y)$.

In fact, both the equations $y = f(x)$ and $x = \varphi(y)$ give the same functional relationship between x and y. Suppose an arbitrary x is given in the direct function. If we mark off an interval from the origin O

along the axis OX, corresponding to the number x, then erect a perpendicular to OX from the end of this interval as far as its intersection with the graph, we obtain the value of y corresponding to the chosen x as the length of this perpendicular, with the corresponding sign. In the case of the inverse function $x = \varphi(y)$, we have only to measure off the given value y from the origin O along the axis OY, a perpendicular to OY then being erected from the end of this segment as far as its intersection with the graph. The length of this perpendicular, with the relevant sign, gives us the value of x corresponding to the chosen y.

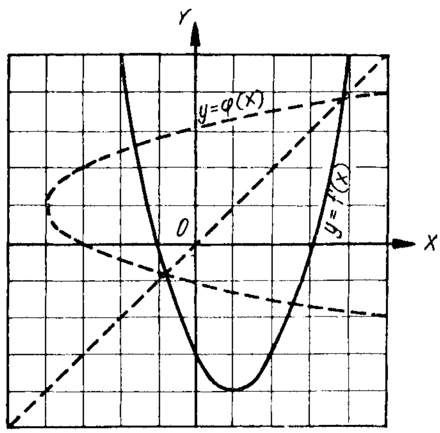

FIG. 25

The inconvenience arises here, that the independent variable x is measured off on one axis, namely OX, in the first case, whilst in the second case the independent variable y is measured off on the other axis, OY. In other words, we can only keep the same graph on transition from the direct function $y = f(x)$ to the inverse function $x = \varphi(y)$, provided we bear in mind that, on making the transition, the axis representing the value of the independent variable becomes the axis representing the function, and vice versa.

To avoid this inconvenience, we turn the plane as a whole on making the transition, so that the axes OX and OY change places. We simply turn the plane of the figure, together with the graph, through 180° about the bisector of the first quadrant of the coordinate axes. The axes having changed places, the inverse function $x = \varphi(y)$ now has to be written in the usual way: $y = \varphi(x)$. Thus, *given the graph of the direct function $y = f(x)$, the graph of the inverse function $y = \varphi(x)$ is obtained simply by turning the plane of the graph through 180° about the bisector of the first quadrant of the coordinate axes.*

The full-line curve of Fig. 25 represents a direct function, the broken curve being its inverse. A broken line is also used for the bisector of the first quadrant of the coordinate axes, the plane of the figure being rotated about this, in order to obtain the broken curve from the full curve.

21. Many-valued functions. It is characteristic of all the graphs of elementary functions considered above, that perpendiculars to the axis OX cut the graph in not more than one, and for the most part in one, point. This means that, given x in the function defined by the graph, one corresponding value of y is defined. A function of this sort is said to be single-valued.

If perpendiculars to OX cut the graph in more than one point this means that, given x, there are several corresponding ordinates of the graph, i.e. several values of y. Such functions are called *many-valued;* they have already been mentioned in [5].

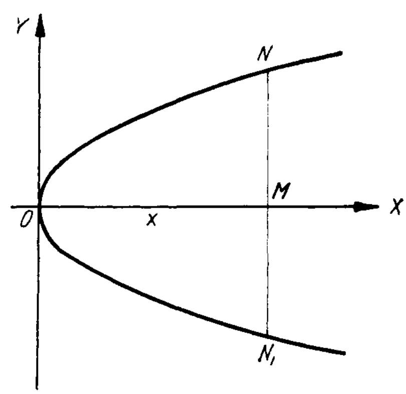

FIG. 26

Although the direct function $y=f(x)$ is single-valued, its inverse $y=\varphi(x)$ can be many-valued. This is evident, for instance, from Fig. 25.

We shall consider in detail one elementary case. The continuous curve of Fig. 13 is the graph of the function $y = x^2$. If the figure is turned through 180° about the bisector of the first quadrant of the axes, the graph of the inverse function $y = \sqrt{x}$ is obtained (Fig. 26).

Let us consider it in detail. For negative x (to the left of axis OY), perpendiculars to OX do not intersect the graph, i.e. the function $y = \sqrt{x}$ is not defined for $x < 0$. This corresponds to the fact that the square root of a negative number has no real value. For positive x, however, perpendiculars to OX cut the graph in two points, i.e. for a given positive x we have two ordinates of the graph: $\overline{MN}$ and $\overline{MN_1}$. The first ordinate gives a certain positive y, the second a negative value of the same absolute magnitude. This corresponds to the fact that the square root of a positive number has two values, equal in absolute magnitude and opposite in sign. It is also evident from the figure that we have only one value $y = 0$ for $x = 0$. Thus, the function $y = \sqrt{x}$, defined for $x \geqslant 0$, has two values for $x > 0$, and one value for $x = 0$.

It may be noted that our function $y = \sqrt{x}$ can be made single-valued by taking only part of the graph of Fig. 26. For instance, let us take only the part in the first quadrant of the axes (Fig. 27). This corresponds to taking only positive values of the square root.

It may also be noted that the part of the graph of the function $y = \sqrt{x}$, shown in Fig. 27, is obtained from the part of the graph of the direct function $y = x^2$ (Fig. 13) lying on the right of axis OY. The part of the graph of

$$y = \sqrt{x} \text{ or } y = x^{\frac{1}{2}}$$

lying in the first quadrant of the axes, has already been illustrated in Fig. 22.

We now turn to the case when rotation of a single-valued direct function leads to a single-valued inverse function. A new concept must be introduced for this.

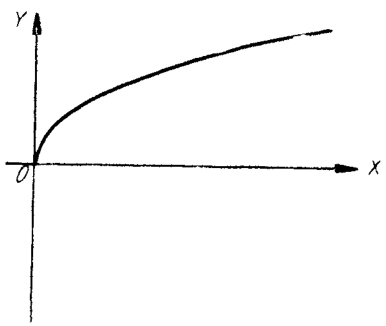

FIG. 27

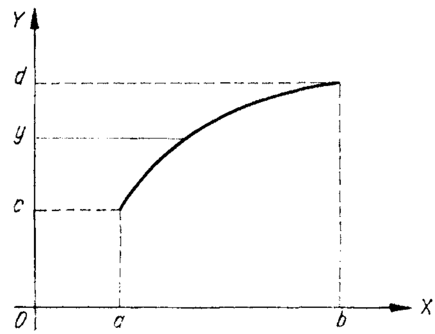

FIG. 28

The function $y = f(x)$ is said to be increasing, if y increases when the independent variable x increases, i.e. if $x_2 > x_1$ implies $f(x_2) > f(x_1)$.

With the axes OX, OY as used by us, increasing x implies movement along OX to the right, and increasing y, movement upwards along OY. It is characteristic of the graph of an increasing function that, on moving along the curve in the direction of increasing x (to the right), we also move in the direction of increasing y (upwards).

Let us consider the graph of a single-valued increasing function, defined in the interval $a \leqslant x \leqslant b$ (Fig. 28). Let $f(a) = c$ and $f(b) = d$, so that evidently, since the function is increasing, $c < d$. If we take any y in the interval $c \leqslant y \leqslant d$, and draw the corresponding perpendicular to the axis OY, this perpendicular will cut our graph in only one point, i.e. for every y in the interval $c \leqslant y \leqslant d$ there is one corresponding definite value of x. In other words, *the inverse of an increasing function is single-valued.*

It is clear from the figure that this inverse function is increasing.

Similarly, *the function $y = f(x)$ is said to be decreasing, if increase of the independent variable x implies decrease of the corresponding y,*

i.e. *if* $x_2 > x_1$ *implies* $f(x_1) > f(x_2)$. It can be shown, as above, that the inverse of a decreasing function is a single-valued, decreasing function. A further important fact will be noted. We assume in all our discussions that *the graph of the function consists of an unbroken, continuous curve*. This is equivalent to a special analytic property of the function $f(x)$, viz, *continuity of the function*. A rigorous mathematical definition of continuity is given in Chapter II, where continuous functions are studied. The present chapter aims merely at a preliminary acquaintance with basic concepts, which are considered systematically in later chapters.

It may be noted as regards terminology, that when no reference is made to a function being many-valued, it can always be assumed to be single-valued.

22. Exponential and logarithmic functions. Let us now return to the study of elementary functions. An exponential function is defined by the equation

$$y = a^x, \tag{15}$$

where the base a is a given positive number, different from unity. The value of a^x is evident for integral positive x. For fractional positive x, a^x is defined as the radical $a^{p/q} = \sqrt[q]{a^p}$, where, with even q, we agree to take the positive value of the radical. Without entering at present into a detailed consideration of the value of a^x for irrational x, it will simply be said that approximate values of a^x are obtained by taking approximate values for the irrational x (as described above in [2]); and the closer the approximation for x, the closer for a^x. For instance, as approximations for $a^{\sqrt{2}}$, where

$$\sqrt{2} = 1.414213\ldots,$$

we have:

$$a^1 = a;\quad a^{1.4} = \sqrt[10]{a^{14}};\quad a^{1.41} = \sqrt[100]{a^{141}};\ldots$$

Evaluation of a^x for negative x follows from evaluation for positive x, since $a^{-x} = 1/a^x$. Since we agreed above always to take the positive radical in the expression $a^{p/q} = \sqrt[q]{a^p}$, it follows that the function a^x is positive for all real x. Apart from this, it may be mentioned in passing that it can be shown that a^x is an increasing function for

$a > 1$, and a decreasing function for $0 < a < 1$. A more detailed study of this function is given later, in [44].

Figure 29 illustrates the graph of function (15) for various values of a. We shall note some peculiarities of these graphs. We have, first of all, $a^0 = 1$ for any a, so that the graph of function (15) passes through the point $y = 1$ on the axis OY for any a, i.e. through the point with coordinates $x=0$, $y = 1$.

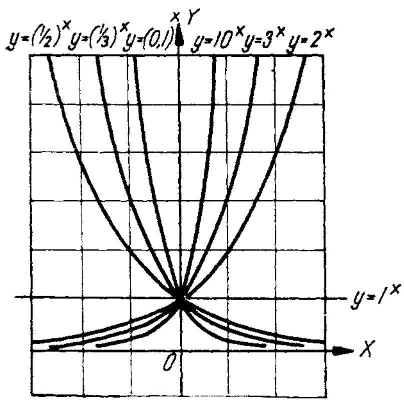

FIG. 29

For $a > 1$, the curve rises indefinitely from left to right (in the direction of increasing x). On moving to the left, it approaches the axis OX indefinitely, without ever touching it. The curve is differently situated relative to the axes for $a < 1$. On moving to the right, the curve indefinitely approaches the axis OX, whilst it rises indefinitely on moving to the left. Since a^x is always positive, the graph is evidently located above the axis OX. It may be noted further, that the graph of the function $y=(1/a)^x$ can be obtained from the graph of $y = a^x$, by turning the figure through 180° about the axis OY. This follows directly from the fact that, with such turning, x becomes $-x$, and $a^{-x} = (1/a)^x$.

A further remark: if $a = 1$, then $y = 1^x$, so that we have $y = 1$ for all x [12].

A *logarithmic function* is defined by the equation:

$$y = \log_a x. \tag{16}$$

By definition of logarithm, function (16) is the inverse of function (15). Thus, we can obtain the graph of the logarithmic function (Fig. 30) from the graph of the exponential, by turning the curves of Fig. 29 through 180° about the bisector of the first quadrant of the axes. Since function (15) is increasing for $a > 1$, the inverse function (16) is also increasing and single-valued, being only defined for $x > 0$, as is evident from Fig. 29 (logarithms of negative numbers do not exist). All the graphs of Fig. 30 cut the axis OX in the point $x = 1$. This corresponds to the fact that the logarithm of unity is zero for any base. For the sake of clarity, Fig. 31 shows a single graph of (16) for $a > 1$.

The concept of *logarithmic scale* and the theory of the *logarithmic slide-rule* are closely associated with the concept of logarithmic function.

A scale is called logarithmic when it is drawn on a given line so that the length of a division, instead of corresponding with the number which denotes

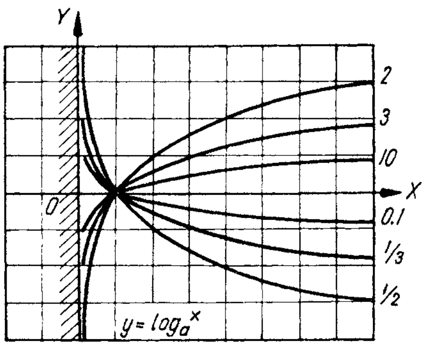

FIG. 30

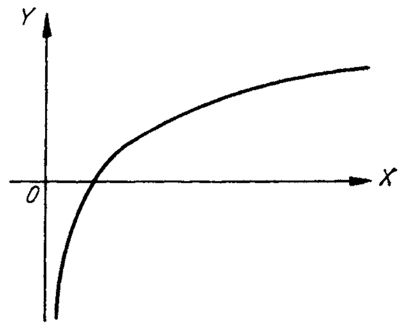

FIG. 31

the division, corresponds with the logarithm of this number, usually to base 10 (Fig. 32). Thus, if a certain division of the scale denotes the number x, the length of the segment $\overline{1x}$ is equal to $\log_{10}x$, instead of x. The length of the segment between two points of the scale, denoting x and y, will be:

$$\overline{1y} - \overline{1x} = \log_{10} y - \log_{10} x = \log_{10} \frac{y}{x} \qquad \text{(Fig. 32);}$$

the logarithm of the product xy is obtained simply by adding segment $\overline{1y}$ to $\overline{1x}$, since the segment thus obtained is equal to:

$$\log_{10} x + \log_{10} y = \log_{10} (xy).$$

Thus, given a logarithmic scale, multiplication and division of numbers can be carried out simply by adding and subtracting segments of the scale, this being realized most simply in practice with the aid of two identical scales, one of which can slide along the other (Figs. 32 and 33). This is the idea underlying the construction of logarithmic slide-rules.

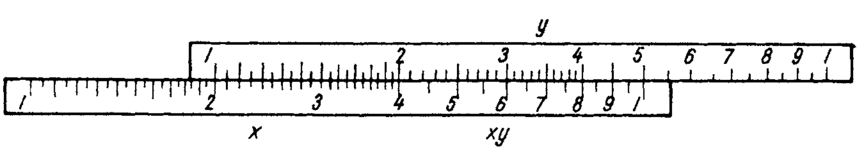

FIG. 32

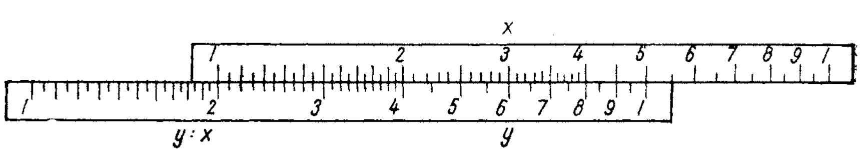

FIG. 33

Logarithmic paper is often used for calculations; this consists of ruled sheets, where the divisions of the axes OX and OY are in accordance with a logarithmic, instead of an ordinary, scale.

23. Trigonometric functions. We shall only consider the four basic trigonometric functions

$$y = \sin x, \quad y = \cos x,$$

$$y = \tan x, \quad y = \cot x,$$

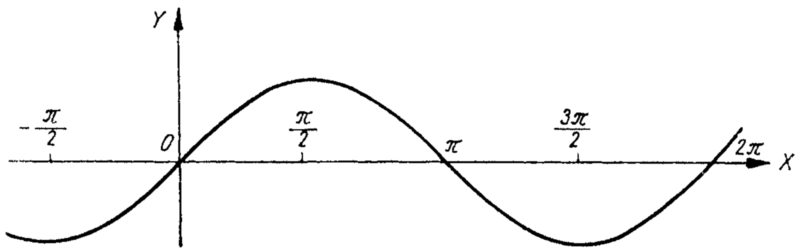

FIG. 34

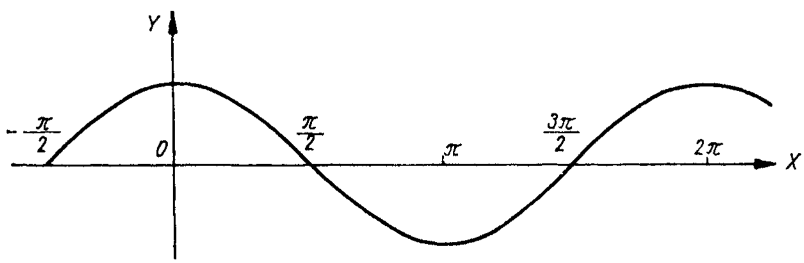

FIG. 35

the independent variable x being measured in radians, i.e. taking as unit angle the angle subtended at the centre of a circle by a segment equal in length to the radius.

The graph of the function $y = \sin x$ is shown in Fig. 34. It is evident from the formula:

$$\cos x = \sin\left(x + \frac{\pi}{2}\right)$$

that the graph of $y = \cos x$ (Fig. 35) can be obtained from the graph of $y = \sin x$ simply by displacing it to the left along the axis OX by an amount $\pi/2$.

The graph of the function $y = \tan x$ is shown in Fig. 36. The curve consists of a series of identical, separate, infinite branches. Each

branch is located in a vertical strip of width π and consists of an increasing function of x. Finally, Fig. 37 gives the graph of $y = \cot x$, which is also made up of separate infinite branches.

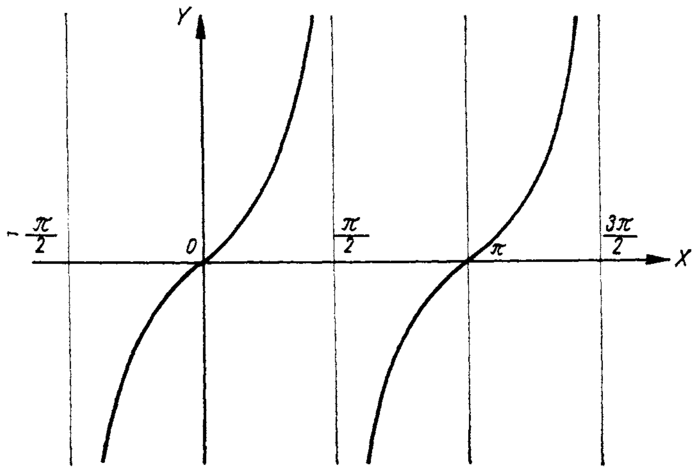

FIG. 36

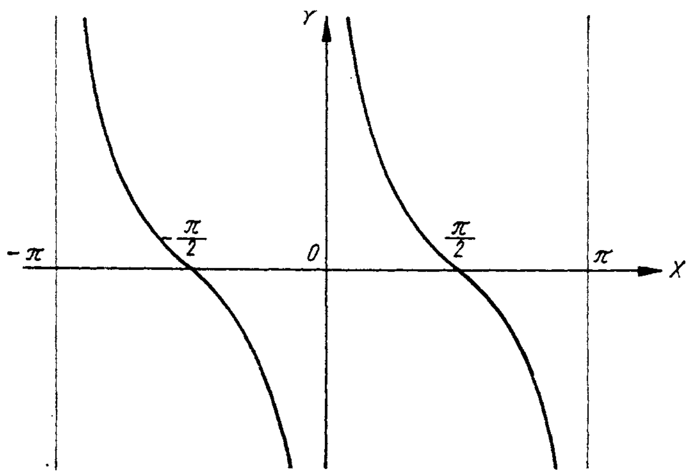

FIG. 37

The graphs obtained by displacing the graphs of $y = \sin x$ and $y = \cos x$ to the left or right along the axis OX by an amount 2π, coincide with the original graphs, corresponding to the fact that functions $\sin x$ and $\cos x$ have period 2π, i.e.

$$\sin(x \pm 2\pi) = \sin x \quad \text{and} \quad \cos(x \pm 2\pi) = \cos x,$$

for any x. The graphs of functions $y = \tan x$ and $y = \cot x$ similarly coincide on displacement along the axis OX by the amount π.

The graphs of the functions:

$$y = A \sin ax,\ y = A \cos ax \ (A > 0,\ a > 0) \tag{17}$$

are always similar to those of $y = \sin x$ and $y = \cos x$. For example, to obtain the graph of the first of functions (17) from the graph of $y = \sin x$, the lengths of all the ordinates of this latter function must be multiplied by A, and the scale on the axis OX changed in such

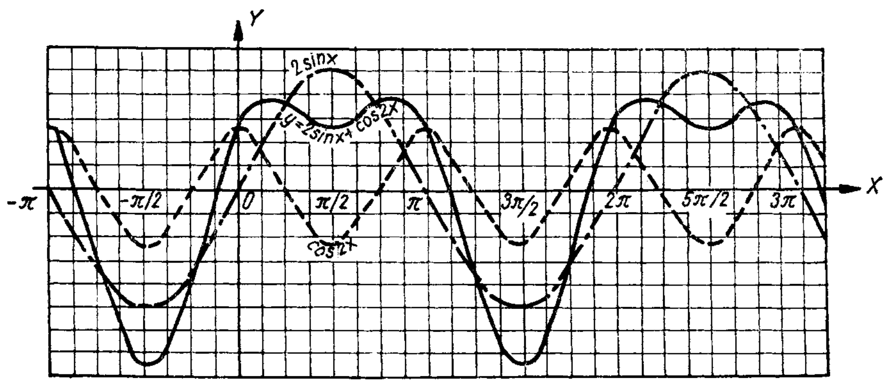

FIG. 38

a way that the point with abscissa x becomes the point with abscissa x/a. Functions (17) are also periodic, with period $2\pi/a$.

The graphs of the more complicated functions:

$$y = A \sin (ax + b),\ y = A \cos (ax + b), \tag{18}$$

which are referred to as *simple harmonic curves*, are obtained from the graphs of functions (17) by displacement to the left along the axis OX by an amount b/a (taking $b > 0$). Functions (18) also have a period of $2\pi/a$.

The graph of another more complicated function:

$$y = A_1 \sin a_1 x + B_1 \cos a_1 x + A_2 \sin a_2 x + B_2 \cos a_2 x,$$

consists of the sum of several terms of type (17), and can be constructed, for example, by adding the ordinates of the graphs of the separate terms. The curves thus obtained are usually referred to as *compound harmonic curves*. Figure 38 illustrates the construction of the graph of the function:

$$y = 2 \sin x + \cos 2x.$$

It may be remarked here, that the function:

$$y = A_1 \sin a_1 x + B_1 \cos a_1 x \tag{19}$$

can be written in the form (18) and represents a simple harmonic oscillation.

We write, in fact:

$$m = \frac{A_1}{\sqrt{A_1^2 + B_1^2}}, \qquad x = \frac{B_1}{\sqrt{A_1^2 + B_1^2}},$$

$$A = \sqrt{A_1^2 + B_1^2}.$$

Obviously:

$$A_1 = mA, \quad B_1 = nA, \tag{20}$$

and also:

$$m^2 + n^2 = 1,$$

$$|m| \leqslant 1, \quad |n| \leqslant 1,$$

so that, from trigonometry, an angle b_1 can always be found such that:

$$\cos b_1 = m, \ \sin b_1 = n. \tag{21}$$

Substituting for A_1 and B_1 in (19) from expressions (20), and using (21), we obtain:

$$y = A(\cos b_1 \cdot \sin a_1 x + \sin b_1 \cdot \cos a_1 x),$$

i.e.

$$y = A \sin (a_1 x + b_1)$$

24. Inverse trigonometric, or circular, functions. These functions are obtained by inversion of the trigonometric functions:

$$y = \sin x, \ \cos x, \ \tan x, \ \cot x,$$

their symbols being respectively:

$$y = \text{arc} \sin x, \ \text{arc} \cos x, \ \text{arc} \tan x, \ \text{arc} \cot x;$$

these symbols are simply abbreviated forms of description for the angle (or arc), of which the sine, cosine, tangent or cotangent is respectively equal to x.

We shall consider the function:

$$y = \text{arc} \sin x. \tag{22}$$

The graph of this function (Fig. 39) is obtained from the graph of $y = \sin x$ by the rule given in [20]. This graph is wholly located in the vertical strip of width two, based on the interval $-1 \leqslant x \leqslant +1$ of the axis OX, i.e. the function (22) is only defined in the interval $-1 \leqslant x \leqslant +1$. Furthermore, equation (22) is equivalent to the equation $\sin y = x$; and, as is known from trigonometry, for a given x we obtain an infinite number of values of y. We see from the graph, in fact, that perpendiculars to the axis OX from points in the interval

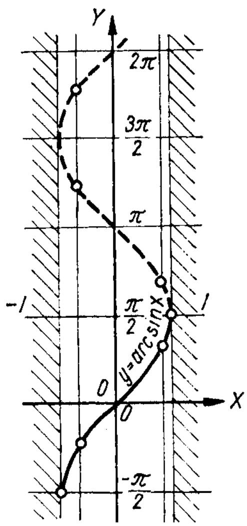

FIG. 39

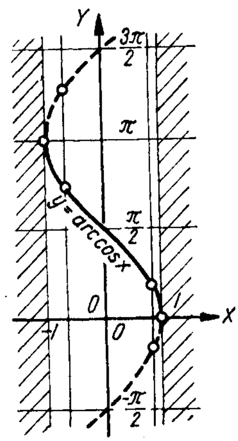

FIG. 40

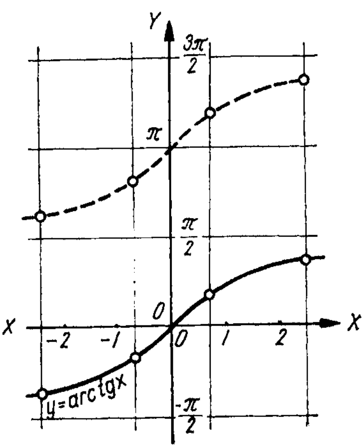

FIG. 41

$-1 \leqslant x \leqslant +1$ have an infinite number of points in common with the graph, i.e. function (22) is many-valued.

We see directly from Fig. 39 that function (22) becomes single-valued if, instead of taking all the graph, we limit ourselves to the part shown in heavier type, which corresponds to stipulating that we shall consider only those values of the angle y, having a given $\sin y = x$, which lie in the interval $(-\pi/2, \pi/2)$.

Figures 40 and 41 illustrate the graphs of $y = \operatorname{arc} \cos x$ and $y = \operatorname{arc} \tan x$, the parts of the graphs in heavier type being those which must be kept in order to make the functions single-valued (we leave it to the reader to draw the figure for $\operatorname{arc} \cot x$). It may be noted here that the functions $y = \operatorname{arc} \tan x$ and $y = \operatorname{arc} \cot x$ are defined for all real values of x.

By noting from the figure the interval of variation of y over the heavier part of the curve, we obtain a table of bounds, within which the function remains single-valued:

y	arc sin x	arc cos x	arc tan x	arc cot x
Inequality for y	$-\frac{\pi}{2} < y < \frac{\pi}{2}$	$0 < y < \pi$	$-\frac{\pi}{2} < y < \frac{\pi}{2}$	$0 < y < \pi$

It can easily be shown that the functions thus defined, called the principal values of the inverse trigonometric functions, satisfy the relationships:

$$\left.\begin{aligned} \operatorname{arc}\ \sin x + \operatorname{arc}\ \cos x &= \frac{\pi}{2} \\ \operatorname{arc}\ \tan x + \operatorname{arc}\ \cot x &= \frac{\pi}{2} \end{aligned}\right\} \tag{23}$$

§ 2. The theory of limits. Continuous functions

25. Ordered variables. When referring to the independent variable x, we have only been concerned with the set of the values that x can assume. For example, this can be the set of values satisfying $0 \leqslant x \leqslant 1$. We shall now consider the variable x taking an infinity of values in sequence, i.e. we are now interested, not only in the set of values of x, but also in the order in which it takes these values. More precisely, we assume the possibility of distinguishing, for every value of x, a value that precedes it and a value that follows it, it being also assumed that no value of the variable is the last, i.e. whatever value we take, there exists an infinity of successive values. A variable of this type is sometimes called *ordered*. If x', x'' are two values of the ordered variable x, a preceding and a succeeding value can be distinguished, whilst if x' precedes x'', and x'' precedes x''', then x' precedes x'''. We shall assume, for example, that the set of values of x is defined by $0 \leqslant x < 1$, and that of two distinct values x' and x'' the succeeding value is the greater. We thus obtain an ordered variable, continuously increasing through all real values from zero to unity, without reaching unity. The sequence of values of the variable, for phenomena occurring in time, is established by the temporal sequence, and we shall sometimes make use of this time-scheme below,

using terms such as "previous" and "later" in place of "preceding" and "succeeding" values.

An important particular case of an ordered variable is that when the sequence of values of the variable can be enumerated, by arranging them in a series of the form:

$$x_1, x_2, x_3 \ldots x_n \ldots$$

so that, given two values x_p' and x_q, the value succeeds that has the greater subscript. In the case mentioned above, when the variable increases from zero to unity, we can clearly not enumerate its successive values. It may also be noted that it is possible to encounter identical values amongst those of an ordered variable. For example, we might have $x_3 = 7$ and $x_{12} = 7$ in the enumerated variable. Abstracting, as we always do, from the concrete nature of the magnitude (length, weight etc.), we must understand by the term "ordered variable", or as we shall say for brevity, "variable", simply the total sequence of its numerical values. We normally introduce one letter, say x, and suppose that it assumes successively the above-mentioned numerical values.

For every value of the variable x, a corresponding point K is defined on the axis OX. Thus, as x varies in sequence, the point K moves along OX.

The present paragraph is devoted to the basic theory of limits, which is fundamental to all modern mathematical analysis. This theory considers some extremely simple, and at the same time, extremely important, cases of variation of magnitudes.

26. Infinitesimals. We assume that the point K constantly remains inside a certain interval of the axis OX. This is equivalent to the condition that the length of the interval $\overline{OK}$, where O is the origin, remains less than a definite positive number M. The magnitude x is said to be *bounded* in this case. Noting that the length of $\overline{OK}$ is $|x|$, we can give the following definition:

DEFINITION. *A variable x is said to be bounded, if there exists a positive number M, such that $|x| < M$ for all values of x.*

We can take $x = \sin a$ as an example of a bounded magnitude, where the angle a varies in any manner. Here, M can be taken as any number greater than unity.

We now consider the case when the point K is displaced successively, and indefinitely approaches the origin. More precisely, we suppose

that successive displacements of point K bring it inside any previously assigned small section $\overline{S'S}$ of the axis OX with centre O, and that it remains inside this section on further displacement. In this case, we say that the *magnitude x tends to zero* or is an *infinitesimal.*

We denote the length of the interval $\overline{S'S}$ by 2ε, where ε signifies any given positive number. If the point K is inside $\overline{S'S}$, then $\overline{OK} < \varepsilon$ and conversely, if $\overline{OK} < \varepsilon$, K is inside $\overline{S'S}$. We can thus give the following definition: *The variable x tends to zero or is an infinitesimal, if for any given positive ε there exists a value of x, such that for all subsequent values of x, $|x| < \varepsilon$.*

In view of the importance of the concept of infinitesimal, we give another formulation of the same definition.

DEFINITION. *A magnitude x is said to tend to zero or to be an infinitesimal, if on successive variation $|x|$ becomes, and on further variation remains, less than any previously assigned small positive number ε.*

The term "infinitesimal" denotes the character of the variation of the variable described above, and the underlying concept is not to be confused with that of a *very small magnitude*, which is often employed in practice.

Suppose that, in measuring a certain tract of land, we obtained 1000 m, with some remainder that we considered very small in comparison with the total length, so that we neglected it. The length of this remainder is expressed by a definite positive number, and the term "infinitesimal" is evidently not applicable here. If we were to meet with the same remainder in a second, more accurate measurement, we should cease to consider it as very small, and we should take it into account. It is thus clear that the concept of a small magnitude is a relative concept, bound up with the practical nature of the measurement.

Suppose that the successive values of the variable x are

$$x_1, x_2, x_3, \ldots x_n, \ldots;$$

and let ε be any given positive number. To prove that x is an infinitesimal, we must show that, starting with a certain value of n, $|x_n|$ will be less than ε, i.e. we must be able to find a certain integer N such that

$$|x_n| < \varepsilon \text{ for } n > N.$$

This N depends on ε.

As an example of an infinitesimal, we take the magnitude assuming successively the values:

$$q, q^2, q^3, \dots q^n, \dots \quad (0 < q < 1). \tag{1}$$

We have to satisfy the inequality:

$$q^n < \varepsilon \text{ or } n \log_{10} q < \log_{10} \varepsilon.$$

Remembering that $\log_{10} q$ is negative, we can rewrite the above inequality as:

$$n > \frac{\log_{10} \varepsilon}{\log_{10} q},$$

since division by a negative number changes the sense of the inequality; thus we can now take N as the largest integer in the quotient $\log_{10} \varepsilon \,/\, \log_{10} q$. Thus the magnitude in question, or as we usually say, the *sequence* (1) tends to zero.

If we replace q by $(-q)$ in the sequence (1), the only difference is the appearance of the minus sign with odd powers; the absolute magnitude of the members of the sequence is as before, and hence we also have an infinitesimal in this case.

The fact that x is infinitesimal is usually denoted by:

$$\lim x = 0 \text{ or } x \to 0.$$

Here, lim is an abbreviation of "limit"

We note two properties of infinitesimals.

1. *The sum of any (definite) number of infinitesimals is also an infinitesimal.*

Take, for example, the sum $w = x + y + z$ of three infinitesimals, and suppose that the variables are enumerated. Let

$$x_1, x_2 \dots;\ y_1, y_2 \dots;\ z_1, z_2 \dots$$

be the successive values of x, y, z, respectively. We obtain successive values for w:

$$w_1 = x_1 + y_1 + z_1,\ w_2 = x_2 + y_2 + z_2, \dots .$$

Let ε be any given positive number. Since x, y, z are infinitesimals, we can say that there exists N_1, such that $|x_n| < \varepsilon/3$ for $n > N_1$; N_2, such that $|y_n| < \varepsilon/3$ for $n > N_2$; and N_3, such that $|z_n| < \varepsilon/3$ for $n > N_3$. If N denotes the greatest of N_1, N_2 and N_3, we have:

$$|x_n| < \frac{\varepsilon}{3};\quad |y_n| < \frac{\varepsilon}{3};\quad |z_n| < \frac{\varepsilon}{3} \text{ for } n > N,$$

and hence:

$$|w_n| \leqslant |x_n| + |y_n + |z_n| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \text{ for } n > N,$$

i.e. $|w_n| < \varepsilon$ for $n > N$, whence $w = x + y + z$ is an infinitesimal. In the general case of non-enumerated variables we can look on x, y, z as functions of some ordered variable t: $x = x(t)$, $y = y(t)$, $z = z(t)$. Variables x, y, z are themselves ordered, so that if $t = t'$ precedes $t = t''$, then $x(t')$ precedes $x(t'')$, etc. The sum

$$w(t) = x(t) + y(t) + z(t) ,$$

obtained by adding the x, y, and z corresponding to the same value of t, is also ordered. The proof is as above, for enumerated variables. In this latter case, t has the role of subscript; or the subscript can be looked on as an increasing, integral t.

2. *The product of a bounded magnitude and an infinitesimal is an infinitesimal.*

We consider the product of the enumerated variables xy, where x is bounded, and y is an infinitesimal. We have the condition that $|x|$ remains less than some positive M for any n. If ε is any given positive number, there exists N, such that $|y_n| < \varepsilon/M$ for $n > N$. Thus

$$|x_n y_n| = |x_n| \cdot |y_n| < M . \frac{\varepsilon}{M} \text{ for } n > N.$$

Hence, $|x_n y_n| < \varepsilon$ for $n > N$, so that $xy \to 0$. The proof is analogous for non-enumerated variables.

We note that the second property is all the more readily justified if x is a constant. We can now take M as any positive number greater than $|x|$, i.e. *the product of a constant and an infinitesimal is an infinitesimal.*

In view of the fundamental importance of the concept of infinitesimal for what follows, we shall pause to add some remarks supplementary to the above.

As we have shown, a variable having the sequence of values (1), tends to zero, only if $0 < q < 1$ or $-1 < q < 0$. Setting $q = 1/2$, for example, we obtain the sequence:

$$\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \ldots \frac{1}{2^n}, \ldots .$$

Each successive value is less than the previous one in this case, and the variable tends to zero, diminishing all the time. Setting $q = -1/2$, we obtain the sequence:

$$-\frac{1}{2}, \frac{1}{4}, -\frac{1}{8}, \frac{1}{16}, \dots$$

Here the variable tends to zero, taking values in turn greater than, and less than, zero.

Suppose that we insert zero in every other place in the above sequence, i. e. we take a variable with the sequence:

$$\frac{1}{2}, 0, \frac{1}{4}, 0, \frac{1}{8}, 0, \frac{1}{26}, 0, \frac{1}{32}, 0, \frac{1}{64}, 0\dots$$

Clearly, the variable in this case tends to zero, though in the process it takes exactly the value zero an infinite number of times. This does not contradict the definition of a magnitude tending to zero.

Finally, suppose that all the successive values of a variable are equal to zero. This also comes under the definition of a magnitude tending to zero, all the more since $|x|$ is now zero all the time, i.e. $|x| < \varepsilon$ for any given positive ε, not only from a certain initial point of its variation, but always. In other words, a constant equal to zero comes under the definition of an infinitesimal. No other constant whatever comes under the definition.

There is one further point. We recall the definition of infinitesimal: for any given positive ε, there exists a value of the variable x, such that for all subsequent values, $|x| < \varepsilon$. It follows immediately, that in proving that a given variable x tends to zero, we can confine ourselves to considering only those values of x that succeed a certain definite value of x, where this definite value can be chosen arbitrarily.

Concerning this, it is useful in the theory of limits to add a rider to the definition of a bounded magnitude, viz, there is no need to demand that $|y| < M$ for all values of y; it is sufficient to take the more general definition: *a magnitude y is said to be bounded, if there exists a positive number M and a value of y, such that $|y| < M$ for all subsequent values.*

The proof of the second property of infinitesimals remains unchanged with this definition of a bounded magnitude. For an enumerated variable, the first definition of a bounded magnitude follows from the second, so that the second is not less

general. In fact, if $|x_n| < M$ for $n > N$, then denoting by M' the greatest of numbers

$$|x_1|, |x_2|, \ldots |x_N| \text{ and } M,$$

we can assert that $|x_n| < M' + 1$ for any n.

27. The limit of a variable. We have called a variable an infinitesimal, if its corresponding point K in the axis OX has on displacement the following property: on successive variation the length of the interval $\overline{OK}$ becomes, and on further variation remains, less than any given positive number ε. We now suppose that this property is fulfilled, not by the interval $\overline{OK}$, but by $\overline{AK}$, where A

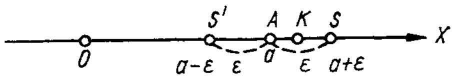

FIG. 42

is a definite point on the axis OX with abscissa a (Fig. 42). In this case, the interval $\overline{S'S}$ of length 2ε will have its centre at the point A, abscissa $x = a$, instead of at the origin, and the point K must come within this interval on successive displacement, then remain there on further displacement. We say in this case that the constant a is *the limit of variable x*, or that *x tends to a.*

Noting that the length of $\overline{AK}$ is $|a - x|$ [9], we can formulate the following definition:

DEFINITION. *The constant a is called the limit of the variable x, when the difference $a - x$ (or $x - a$) is an infinitesimal.*

Having regard to the definition of an infinitesimal, a limit can be thus defined:

DEFINITION. *The constant a is called the limit of the variable x, when we have the following property: for any given positive ε there exists a value of x such that, for all subsequent values, $|a - x| < \varepsilon$.*

We note some immediately obvious consequences of this definition, without dwelling on their detailed proof.

No variable can tend to two different limits, and not every variable has a limit. For example, the variable $\sin \alpha$ oscillates between -1 and 1 on successive increase of the angle α, and has no limit.

The limit of an infinitesimal is zero.

If x and y vary simultaneously, and each tends to a limit in the course of successive variation, whilst both always satisfy $x \leqslant y$, their limits a and b satisfy the condition $a \leqslant b$.

We note here, that if the variables satisfy $x < y$, the sign of equality can be obtained for their limits, i.e. we have $a \leqslant b$.

If x, y, z vary simultaneously and always satisfy the condition $x \leqslant y \leqslant z$ on successive variation, and if x and z tend to the same limit a, y also tends to the limit a.

If a is the limit of x (or x tends to a), we write:

$$\lim x = a, \text{ or } x \to a.$$

If x tends to a, the difference $x - a = \alpha$ is an infinitesimal, and we can write:

$$x = a + \alpha \tag{2}$$

i.e. *every variable tending to a limit can be expressed as the sum of two terms: a constant term, equal to the limit, and an infinitesimal. Conversely, if a variable x can be expressed in the form* (2), *where a is a constant, and α is an infinitesimal, the difference $x - a$ will be an infinitesimal, and hence, a is the limit of x.*

If the sequence $x_1, x_2, \ldots$ tends to the limit a, every infinite subsequence $x_{n_1}, x_{n_2}, \ldots$ contained in the first sequence, also tends to a. In this subsequence, the subscript n_k increases with increasing k and runs through some part of the set of positive integers. There is no analogous property, generally speaking, for a non-enumerated variable tending to a limit.

We take as an example the variable x with the sequence of values:

$$x_1 = 0.1, \; x_2 = 0.11, \; x_3 = 0.111, \; \ldots$$

$$x_n = \overbrace{0.11\ldots11}^{n}, \; \ldots,$$

and we show that its limit is $1/9$. We first form the difference $1/9 - x_n$:

$$\frac{1}{9} - x_1 = \frac{1}{90}, \quad \frac{1}{9} - x_2 = \frac{1}{900}, \quad \frac{1}{9} - x_3 = \frac{1}{9000}, \ldots,$$

$$\frac{1}{9} - x_n = \frac{1}{9 \cdot 10^n}.$$

The condition:

$$\frac{1}{9 \cdot 10^n} < \varepsilon$$

is evidently equivalent to the condition:

$$9 \times 10^n > \frac{1}{\varepsilon} \text{ or } n > \log_{10}\frac{1}{\varepsilon} - \log_{10} 9,$$

and we can take N as the greatest integer contained in the difference $\log_{10} 1/\varepsilon - \log_{10} 9$. In this example, the difference $1/9 - x_n$ is a positive number for every n, i. e. x tends to the limit $1/9$ whilst always remaining less than it.

We now consider the sum of the first n members of the indefinitely diminishing geometrical progression:

$$s_n = b + bq + bq^2 + \ldots + bq^{n-1} \quad (0 < |q| < 1).$$

As we know,

$$s_n = \frac{b(1-q^n)}{1-q}.$$

Setting $n = 1, 2, 3, \ldots$, we obtain the sequence:

$$s_1, s_2, s_3, \ldots s_k, \ldots .$$

We have from the expression for s_n:

$$\frac{b}{1-q} - s_n = \frac{bq^n}{1-q}.$$

The right-hand side consists of the product of a constant $b/(1-q)$ and an infinitesimal q^n [26]. Hence, using the second property of infinitesimals [26], the difference $b/(1-q) - s_n$ is an infinitesimal, and we can say that the constant $b/(1-q)$ is the limit of the sequence $s_1, s_2, \ldots s_k$.

Suppose that $b > 0$ and $q < 0$. The difference $b/(1-q) - s_n$ is now positive for even n and negative for odd n, so that the variable s_n is alternately greater than, and less than, the limit to which it tends.

The same remarks apply in the case of magnitudes that tend to a given limit as were made in the previous paragraph, *apropos* magnitudes that tend to zero.

Any constant, equal to the number a, comes under the definition of a variable, tending to the limit a. We note here, that a magnitude, all of whose values are equal to a, has in the ordinary way an infinite set of values, though all these values are equal to the same number. This view of a constant as a particular case of a variable comes in useful later on.

Furthermore, there is no need to consider all the values of a variable x when defining its limit; we need only take values subsequent to some arbitrarily given value.

Another point: if a variable x tends to a limit a, it will differ from a by as little as is desired, after a certain initial moment of its variation, and hence it is all the more a bounded variable.

An ordered variable does not always have a limit, as already mentioned. If we take, for example, the enumerated variable $x_1 = 0.1$; $x_2 = 0.11$; $x_3 = 0.111, \ldots$, whose limit is $1/9$, and the variable $y_1 = 1/2$; $y_2 = 1/2^2$; $y_3 = 1/2^3, \ldots$, whose limit is zero, the enumerated variable $z_1 = 0.1$; $z_2 = 1/2$; $z_3 = 0.11$; $z_4 = 1/2^2$; $z_5 = 0.111$; $z_6 = 1/2^3; \ldots$, does not tend to a limit. The sequence of its values $z_1, z_3, z_5 \ldots$ has the limit $1/9$, and the sequence $z_2, z_4, z_6 \ldots$ has the limit zero.

28. Basic theorems. 1. *The limit of the algebraic sum of a finite number of variables is equal to the sum of their limits.*

For the sake of exactness let us take the algebraic sum $x - y + z$ of three simultaneously varying magnitudes. We suppose that x, y and z tend respectively to limits a, b and c. We show that the sum tends to the limit $a - b + c$.

We have by hypothesis [27]:

$$x = a + \alpha, \; y = b + \beta, \; z = c + \gamma\,,$$

where α, β, γ are infinitesimals. We can write for the sum:

$$x - y + z = (a + \alpha) - (b + \beta) + (c + \gamma) =$$
$$= (a - b + c) + (\alpha - \beta + \gamma).$$

The first bracket on the right-hand side of this equation is a constant, and the second is an infinitesimal [26]. Hence:

$$\lim (x - y + z) = a - b + c = \lim x - \lim y + \lim z.$$

2. *The limit of the product of a finite number of variables is equal to the product of their limits.*

We confine ourselves to the case of the product xy of two variables. We suppose that x and y vary simultaneously, tending respectively to limits a and b, and we show that xy tends to the limit ab.

We have by hypothesis:

$$x = a + \alpha, \; y = b + \beta\,,$$

where α and β are infinitesimals; hence:

$$xy = (a + \alpha)(b + \beta) = ab + (a\beta + b\alpha + \alpha\beta).$$

Using both of the properties of infinitesimals from [26], we see that the sum in the bracket on the right of this equation is an infinitesimal, and hence we have:

$$\lim (xy) = ab = \lim x \cdot \lim y.$$

3. *The limit of a quotient is equal to the quotient of the limits, provided the limit of the denominator is not zero.*

We take the quotient x/y, and suppose that x and y tend simultaneously to their respective limits a and b, where $b \neq 0$. We show that x/y tends to a/b.

To prove the theorem, it is sufficient to show that the difference $a/b - x/y$ is an infinitesimal. By hypothesis:

$$x = a + \alpha;\; y = b + \beta \qquad (b \neq 0),$$

where α and β are infinitesimals. Hence:

$$\frac{a}{b} - \frac{x}{y} = \frac{1}{b(b+\beta)} \cdot (a\beta - b\alpha).$$

The denominator of the fraction on the right of this equation is the product of two factors, and tends to b^2. Thus, from some initial moment of its variation, it is greater than $b^2/2$, the fraction as a whole being included between zero and $2/b^2$, i.e. the fraction is bounded. The term $(a\beta - b\alpha)$ is an infinitesimal. Hence [26], the difference $a/b - x/y$ is an infinitesimal, and

$$\lim \frac{x}{y} = \frac{a}{b} = \frac{\lim x}{\lim y}.$$

The theorems proved are of fundamental importance in the theory of limits. The proofs have been given for the general case, and not for the case of enumerated variables, as when proving the properties of infinitesimals. But the remark we made when proving the first property of infinitesimals should be borne in mind. Take the case of a product. We take x and y as functions of some ordered variable $t : x = x(t);\ y = y(t)$. Then x and y are themselves ordered variables. The same can be said of their product: $w(t) = x(t) \cdot y(t)$. The subscript plays the part of t in enumerated variables, increasing through integral values.

We remark further, that the above theorems establish the existence of the limit of a sum, a product and a fraction. For example, the third

theorem can be stated more fully as: if numerator and denominator tend to limits, and the limit of the denominator differs from zero, the quotient then tends to a limit, and this limit is the quotient of the limits of numerator and denominator.

We note some consequences of these theorems. If x tends to the limit a, then bx^k, where b is a constant and k a positive integer, tends to the limit ba^k, in accordance with Theorem 2.

Consider the integral polynomial

$$f(x) = a_0 x^m + a_1 x^{m-1} + \ldots + a_k x^{m-k} + \ldots + a_{m-1} x + a_m,$$

with constant coefficients a_k. Using Theorem 1 and the previous remark, we can say that this polynomial tends to a limit:

$$\lim f(x) = f(a) \text{ as } x \text{ tends to } a \text{ where } f(a) = a_0 a^m + a_1 a^{m-1} \ldots \\ \ldots + a_k a^{m-k} + \ldots a_{m-1} a + a_m. \tag{3}$$

Similarly, as x tends to a, the rational fraction:

$$\varphi(x) = \frac{a_0 x^m + a_1 x^{m-1} + \cdots + a_{m-1} x + a_m}{b_0 x^p + b_1 x^{p-1} + \cdots + b_{p-1} x + b_p}$$

tends to a limit:

$$\lim \varphi(x) = \varphi(a) = \frac{a_0 a^m + a_1 a^{m-1} \cdots + a_{m-1} a + a_m}{b_0 a^p + b_1 a^{p-1} \cdots + b_{p-1} a + b_p}, \tag{4}$$

$$\text{if } b_0 a^p + b_1 a^{p-1} \ldots + b_{p-1} a + b_p \neq 0.$$

All these remarks are valid, in whatever way x tends to its limit a.

We can of course take polynomials arranged in powers of several variables, all tending to limits, instead of polynomials arranged in powers of a single variable.

For example, if $\lim x = a$ and $\lim y = b$, then $\lim (x^2 + xy + y^2) = a^2 + ab + b^2$.

29. Infinitely large magnitudes. If the variable x tends to a limit, it is evidently bounded, as already remarked. We now consider some cases of variation of unbounded magnitudes.

As before, we shall take along with x its corresponding point K, displaced on the axis OX. Let the point K move in such a way that, however large an interval $\overline{T'T}$ we take, with the origin as centre, the point K will eventually be displaced outside it, and from then on will remain outside. In this case, x is an infinitely large magnitude, and tends to infinity. Let $2M$ be the length of the interval $\overline{T'T}$. Recalling

that the length of the interval $\overline{OK} = |x|$, we can give the following definition:

The magnitude x is said to be infinitely large, or to tend to infinity, if on successive variation of x, $|x|$ becomes, and on further variation remains, greater than any given positive number M. In other words, the magnitude x is called infinitely large if it satisfies the following condition: *given any positive number M, there exists a value of x such that, for all subsequent x, $|x| > M$.*

In particular, if x is infinitely large, and always remains positive during its successive variation as from a certain initial value (point K to the right of O), we say that x tends to plus infinity ($+\infty$). Similarly, if x remains negative (point K to the left of O), we say that x tends to minus infinity ($-\infty$).

The following symbols are used for infinitely large magnitudes:

$$\lim x = \infty, \ \lim x = +\infty, \ \lim x = -\infty,$$

or

$$x \to \infty \qquad x \to +\infty \qquad x \to -\infty.$$

The term "infinitely large" serves merely as a brief designation for the character of variation described above of the variable x, and here, as with the concept of infinitesimal, a distinction must be made between the concepts of "infinitely large" and "very large" magnitudes.

If, for example, x takes the sequence of values $1, 2, 3, \ldots$ then evidently, $x \to +\infty$. If its sequence of values is: $-1, -2, -3, \ldots$, then $x \to -\infty$. And finally, if the sequence is: $-1, 2, -3, 4, \ldots$, we can write: $x \to \infty$.

Let us take as a further example the magnitude with the sequence of values:

$$q, q^2, q^3, \ldots q^n, \ldots, \ (q > 1), \tag{5}$$

and let M be any given positive number. The condition

$$q^n > M$$

is equivalent to

$$n > \frac{\log_{10} M}{\log_{10} q},$$

and hence, if N is the greatest integer contained in the quotient $\log_{10} M \cdot \log_{10} q$, we have:

$$q^n > M \text{ for } n > N,$$

i.e. the variable in question tends to $+\infty$.

If q is replaced by $(-q)$ in the sequence (5), the only change is in the signs of odd powers of q, the absolute values of the members of the sequence remaining as before; thus, for negative q, with absolute value greater than unity, the sequence (5) tends to infinity.

When in future we say that a variable tends to a limit, a finite limit is to be understood. It is occasionally said that a variable "tends to an infinite limit", implying by these words an infinitely large magnitude.

An immediate consequence of the above definitions is: if variable x tends to zero, then m/x, where m is a given constant, differing from zero, tends to infinity; and if x tends to infinity, m/x tends to zero.

30. Monotonic variables. The important thing is often to show that a given variable tends to a limit, without necessarily being able to discover what this limit actually is. We now outline an important test for the existence of a limit.

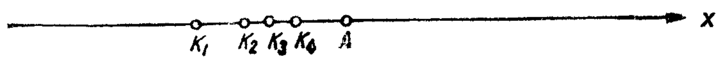

FIG. 43

We suppose that the variable x is always increasing (more precisely never decreasing) or else always decreasing (more precisely, never increasing). In the first case, any given value is not less than all preceding values, and not greater than all subsequent values. In the second case, any given value is not greater than all preceding, and not less than all succeeding, values. We speak of *monotonic variation* in these cases.

Point K on the axis OX, corresponding to x, is now displaced in a single direction, positively, if x increases, and negatively, if x decreases. It is obvious at once that only two possibilities can arise: either K moves away indefinitely along the line ($x \to +\infty$ or $-\infty$), or K indefinitely approaches some definite point A (Fig. 43), i.e. x tends to a limit. If x is known to be bounded, as well as varying monotonically, the first possibility drops out, and it can be asserted that the variable tends to a limit.

This argument is based on intuition, and evidently lacks the force of a proof. We shall give the rigorous proof later.

The above test for the existence of a limit is usually formulated as follows: *if a variable is bounded and varies monotonically, it tends to a limit.*

Take the example of the sequence:

$$u_1 = \frac{x}{1}, \quad u_2 = \frac{x^2}{2!}, \quad u_3 = \frac{x^3}{3!}, \ldots u_n = \frac{x^n}{n!} \ldots, † \qquad (6)$$

where x is a given positive number.

We have:

$$u_n = u_{n-1} \frac{x}{n}. \qquad (7)$$

For $n > x$, x/n is less than unity, and $u_n < u_{n-1}$, i.e. from some initial value, u_n is always decreasing for n increasing, whilst remaining greater than zero. The variable thus tends to some limit u, in accordance with the test for the existence of a limit. Let the integer n increase indefinitely in equation (7). We obtain in the limit:

$$u = u \cdot 0 \text{ or } u = 0,$$

i.e.

$$\lim_{n \to +\infty} \frac{x^n}{n!} = 0. \qquad (8)$$

If we replace x by $(-x)$ in sequence (6), the only change is in the sign of members with odd n, so that the new sequence also tends to zero, i.e. equation (8) is valid for any given x, positive or negative.

We obtain the limit in this example, after first showing that it exists. If we did not show its existence, our method could lead to a false result. Consider, for instance, the sequence:

$$u_1 = q, \ u_2 = q^2, \ \ldots \ u_n = q^n, \ \ldots \ (q > 1).$$

We have obviously:

$$u_n = u_{n-1} q.$$

We denote the limit of u_n by u, without troubling about its existence. On transition to the limit in the above equation, we obtain:

$$u = uq, \text{ i.e. } u(1 - q) = 0,$$

and hence,

$$u = 0.$$

But this result is false, since we know that for $q > 1$, $\lim q^n = + \infty$ [29].

31. Cauchy's test for the existence of a limit. The test for the existence of a limit given in [30] is a sufficient, but not a necessary

† The symbol $n!$ called "factorial n", is short for the product $1 \cdot 2 \cdot 3 \ldots n$.

condition that a limit exists, since, as we know [27], a variable can tend to a limit without varying monotonically.

The French mathematician Cauchy gave a necessary and sufficient condition for the existence of a limit, which we shall now formulate. If the limit is known, it is characterized by the fact that, starting with a certain value of the variable, the absolute value of the difference between the limit and the variable is less than any given positive ε. According to Cauchy's test, a necessary and sufficient condition for a limit to exist is that, starting from a certain value of the variable, the difference between any two successive values of the variable is less than any given positive ε. We formulate this rigorously:

CAUCHY'S TEST. *A necessary and sufficient condition for a variable x to have a limit is that, given any positive number ε, there exists a value of x such that, for any two successive values x' and x'', we have $|x' - x''| < \varepsilon$.*

Suppose that we have the enumerated variable

$$x_1, x_2, \ldots x_n, \ldots$$

According to Cauchy's test, a necessary and sufficient condition for this sequence to have a limit is that, given any positive ε, there exists an N (depending on ε) such that

$$|x_m - x_n| < \varepsilon, \quad \text{for } m \text{ and } n > N. \tag{9}$$

It is easy to show that this condition is necessary. If our sequence has the limit a, we write $x_m - x_n = (x_m - a) + (a - x_n)$, whence it follows:

$$|x_m - x_n| \leqslant |x_m - a| + |a - x_n|.$$

But, by definition of a limit, there exists N such that $|x_m - a| < \varepsilon/2$ and $|a - x_n| < \varepsilon/2$ for m and $n > N$, and therefore $|x_m - x_n| < \varepsilon$

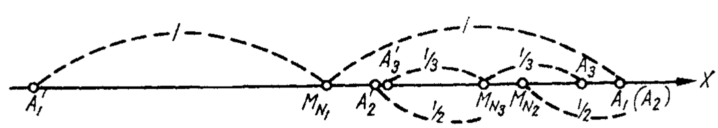

FIG. 44

for m and $n > N$. To put the matter briefly, values of x lying arbitrarily close to a lie arbitrarily close to each other.

We avoid a rigorous proof at present of the sufficiency of Cauchy's test and give a descriptive explanation instead (Fig. 44).

Let M_s be a point of the coordinate axis corresponding to the number x_s. Suppose that condition (9) is fulfilled. In accordance with this condition, there exists a value $N = N_1$ such that

$$|x_s - x_{N_1}| < 1,$$

for $s > N_1$, i.e. every point M_s, where $s > N_1$, lies inside the interval $\overline{A_1'A_1}$, the length of which is equal to two and the mid-point of which corresponds to x_{N_1}.

Similarly, there exists a value $N = N_2 \geqslant N_1$, such that

$$|x_s - x_{N_2}| < \frac{1}{2} \text{ for } s > N_2.$$

We construct an interval, of length unity, with mid-point M_{N_2}; and we let $\overline{A_2'A_2}$ be the part of this interval belonging to $\overline{A_1'A_1}$. By virtue of the two conditions above, the point M_s must lie inside interval $\overline{A_2' A_2}$ for $s > N_2$.

Similarly there exists $N = N_3 \geqslant N_2$, such that $|x_s - x_{N_3}| < 1/3$ for $s > N_3$. We proceed as before, and construct $\overline{A_3'A_3}$, with length not exceeding 2/3 and belonging to $\overline{A_2'A_2}$, all values of M_s being interior points of it for $s > N_3$. Setting $\varepsilon = 1/4, 1/5, \ldots, 1/n, \ldots$, we obtain in this way a sequence of intervals $\overline{A_n'A_n}$, each successive member of which is comprised in the previous member, whilst the length of the members tends to zero. The ends of these intervals obviously tend to the same point A, and the number a corresponding to this point is the limit of the variable x, since it follows from the construction described above that, for a sufficiently large value of s, all the points M_s will lie as close as desired to the point A.

As an application of Cauchy's test, we take Kepler's equation, which defines the position of a planet in its orbit. This equation has the form:

$$x = q \sin x + a,$$

where a and q are given numbers, both lying between zero and unity, and x is unknown.

We take an arbitrary x_0 and construct a sequence of numbers:

$$x_1 = q \sin x_0 + a, \qquad x_2 = q \sin x_1 + a, \ldots,$$

$$x_n = q \sin x_{n-1} + a, \qquad x_{n+1} = q \sin x_n + a, \ldots$$

Subtracting the first equation from the second term by term, we obtain:

$$x_2 - x_1 = q(\sin x_1 - \sin x_0) = 2q \sin \frac{x_1 - x_0}{2} \cos \frac{x_1 + x_0}{2}.$$

Noting that $|\sin a| \leqslant |a|$ and $|\cos a| \leqslant 1$, we have:

$$|x_2 - x_1| \leqslant 2q \frac{|x_1 - x_0|}{2} = q\,|x_1 - x_0| \tag{10}$$

We can find in precisely the same way that

$$|x_3 - x_2| \leqslant q\,|x_2 - x_1|,$$

so that, using (10), we can write:

$$|x_3 - x_2| \leqslant q^2\,|x_1 - x_0|.$$

Proceeding in this manner, we obtain for every n the condition:

$$|x_{n+1} - x_n| \leqslant q^n\,|x_1 - x_0|. \tag{11}$$

We now consider the difference $x_m - x_n$, taking $m > n$ for the sake of clarity:

$$x_m - x_n = x_m - x_{m-1} + x_{m-1} - x_{m-2} + x_{m-2} - x_{m-3} + \cdots + x_{n+1} - x_n.$$

Using (11), and the formula for the sum of the terms of a geometrical progression, we may write:

$$|x_m - x_n| \leqslant |x_m - x_{m-1}| + |x_{m-1} - x_{m-2}| + |x_{m-2} - x_{m-3}| + \cdots$$
$$\cdots |x_{n+1} - x_n| \leqslant (q^{m-1} + q^{m-2} + q^{m-3} + \cdots + q^n)\,|x_1 - x_0| =$$
$$= q^n \frac{1 - q^{m-n}}{1 - q}\,|x_1 - x_0|.$$

As n tends to infinity, q^n tends to zero [26]; $|x_1 - x_0|$ is constant; the fraction $(1-q^{m-n})/(1-q)$ always lies between zero and $1/(1-q)$, i.e. is bounded, since, for $m > n$, q^{m-n} lies between zero and unity. Thus, with indefinite increase of n, and any $m > n$, the difference $x_m - x_n$ tends to zero, and condition (9) is fulfilled. We can say, in accordance with Cauchy's test, that a limit exists:

$$\lim_{n \to +\infty} x_n = \xi.$$

We let n tend to infinity in the equation

$$x_{n+1} = q \sin x_n + a.$$

Using the continuity of the function $\sin x$,† we find in the limit:

$$\xi = q \sin \xi + a, \tag{12}$$

i.e. the limit ξ of the variable x_n is also the root of Kepler's equation.

We started with an arbitrary x_0 in constructing the sequence x_n. We show, however, that Kepler's equation does not possess two roots, i.e. that $\lim x_n = \xi$ is independent of the choice of x_0 and is equal to the *single* root of Kepler's equation.

† The definition of continuity is given below [34].

We assume there is a root ξ_1 in addition to the root ξ, so that:

$$\xi_1 = q \sin \xi_1 + a.$$

Subtracting equation (12) term by term from this equation, we obtain:

$$\xi_1 - \xi = q(\sin \xi_1 - \sin \xi) = 2q \sin \frac{\xi_1 - \xi}{2} \cos \frac{\xi_1 + \xi}{2},$$

whence, as before,

$$|\xi_1 - \xi| < q |\xi_1 - \xi|.$$

But q lies between zero and unity, so that the above relationship is only possible for $\xi_1 - \xi = 0$, i.e. $\xi_1 = \xi$, and hence Kepler's equation has only one root ξ.

32. Simultaneous variation of two variables, connected by a functional relationship. We consider two variables x and y, connected by the functional relationship:

$$y = f(x),$$

and we let $f(x)$ be defined to the left and right of the point $x = c$. We shall assume that x increases and passes through all real values as it tends to c, without in fact reaching c. In this case, $f(x)$ is an ordered variable. We suppose that it has a limit A.

This is usually written as follows:

$$\lim_{x \to c-0} y = \lim_{x \to c-0} f(x) = A, \tag{13}$$

where the symbol $x \to c - 0$ indicates that x tends to c from the side of lower values.

Similarly, if x tends to c whilst diminishing and passing through all real values, and if $f(x)$ now tends to the limit B, we write this as:

$$\lim_{x \to c+0} y = \lim_{x \to c+0} f(x) = B. \tag{14}$$

The existence of the limit (13) is evidently equivalent to $f(x)$ coming as close as desired to the number A, when x comes sufficiently close to the number c, whilst remaining less than c, i.e. (13) is equivalent to the following: *for any given positive number ε there exists a positive number η such that*

$$|A - f(x)| < \varepsilon \text{ as soon as } 0 < c - x < \eta.$$

Of course, η depends on ε.

In precisely the same way, (14) is equivalent to: *for any given positive number ε there exists a positive number η such that*

$$|B - f(x)| < \varepsilon \text{ as soon as } 0 < x - c < \eta.$$

If limits A and B are equal, we write this as follows:

$$\lim_{x \to c} y = \lim_{x \to c} f(x) = A. \tag{15}$$

It is immaterial here, whether x is on one side of c or the other, and (15) implies: for any given positive ε there exists a positive η such that

$$| A - f(x) | < \varepsilon \text{ as soon as } | c - x | < \eta \text{ and } x \neq c. \tag{16}$$

Limit (13) is often denoted by the symbol $f(x - 0)$ and limit (14) by $f(c + 0)$:

$$\lim_{x \to c-0} f(x) = f(c - 0); \quad \lim_{x \to c+0} f(x) = f(c + 0).$$

Symbols $f(c - 0)$ and $f(c + 0)$ should be distinguished from $f(c)$, i.e. the value of $f(x)$ for $x = c$. This latter value can differ from $f(c - 0)$ and $f(c + 0)$, or in fact can be entirely meaningless. The limits $f(c - 0)$ and $f(c + 0)$ exist in the case of functions having graphs with no discontinuities, when we obviously have: $f(c - 0) = f(c + 0) = f(c)$, i.e.

$$\lim_{x \to c} f(x) = f(c).$$

We say in this case that *the function $f(x)$ is continuous for $x = c$ (at the point $x = c$)*. We shall consider the properties of continuous functions in detail later.

We return to the general case. The above definitions are easily generalized for the case when y tends to infinity. It is easy to see, for example, on the basis of what has been said, that

$$\lim_{x \to c-0} \frac{1}{x - c} = -\infty; \quad \lim_{x \to c+0} \frac{1}{x - c} = +\infty,$$

$$\lim_{x \to \frac{\pi}{2} - 0} \tan x = +\infty; \quad \lim_{x \to \frac{\pi}{2} + 0} \tan x = -\infty$$

Taking the principal values of the function $y = \operatorname{arc} \tan x$ [24], we can write:

$$\lim_{x \to c-0} \operatorname{arc} \tan \frac{1}{x - c} = -\frac{\pi}{2};$$

$$\lim_{x \to c+0} \operatorname{arc} \tan \frac{1}{x - c} = \frac{\pi}{2}.$$

If $f(x)$ is defined for all sufficiently large x, the limit can exist:

$$\lim_{x \to +\infty} f(x) = A.$$

If $f(x)$ is defined for all x, either positive or negative, that are sufficiently large in absolute value, the limit can exist:

$$\lim_{x \to \infty} f(x) = A.$$

The latter is equivalent to: for any given positive number ε there exists a positive number M, such that

$$| A - f(x) | < \varepsilon \text{ for } | x | > M.$$

The following equations may easily be verified:

$$\lim_{x \to +\infty} x^3 = + \infty; \quad \lim_{x \to -\infty} x^3 = - \infty;$$

$$\lim_{x \to \infty} \frac{1}{x} = 0; \quad \lim_{x \to \infty} x^2 = +\infty;$$

$$\lim_{x \to \infty} \frac{2x^2 - 1}{3x^2 + x + 1} = \lim_{x \to \infty} \frac{2 - \frac{1}{x^2}}{3 + \frac{1}{x} + \frac{1}{x^2}} = \frac{2}{3};$$

$$\lim_{x \to \infty} \frac{3x + 5}{x^2 + 1} = \lim_{x \to \infty} \frac{\frac{3}{x} + \frac{5}{x^2}}{1 + \frac{1}{x^2}} = 0.$$

We also take an example from physics. Suppose that we heat a certain solid, and let t_0 be its initial temperature. The temperature of the body rises on heating, until the melting point is reached. The temperature now remains constant on further heating, till the point when the whole of the substance has passed over to the liquid state; after this, the temperature-rise begins again, in the resultant liquid. The situation is similar on passage from the liquid to the gaseous state. We shall consider the amount of heat Q communicated to the substance as a function of the temperature. Figure 45 shows the graph of this function, with temperature on the horizontal axis, and the amount of heat absorbed on the vertical axis. Let t_1 be the temperature at which transition to the liquid state begins, and t_2 the temperature at which the transition from the liquid to the gaseous state begins. Evidently:

$$\lim_{t \to t_1 - 0} Q = \text{ord. } \overline{AB} \text{ and } \lim_{t \to t_1 + 0} Q = \text{ord. } \overline{AC}.$$

The size of the segment $\overline{BC}$ gives the latent heat of fusion, and that of $\overline{EF}$ the latent heat of vaporization.

If limits $f(c-0)$ and $f(c+0)$ exist and differ, their difference $f(c+0)-f(c-0)$ is called the break, or jump, of function $f(x)$ at $x = c$ (at the point $x = c$).

The function $y = \arctan 1/(x-c)$ has a jump of π at $x = c$. The function $Q(t)$ just considered has a jump equal to the latent heat of fusion at the melting-point $t = t_1$.

In defining the limit of $f(x)$ as x tends to c, we assumed that x never actually coincides with c. This proviso is made, since the value

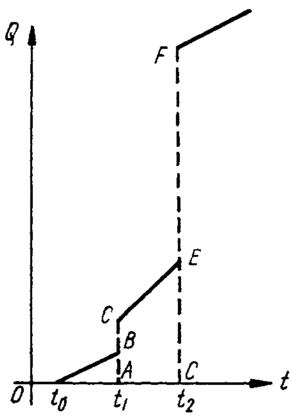

FIG. 45

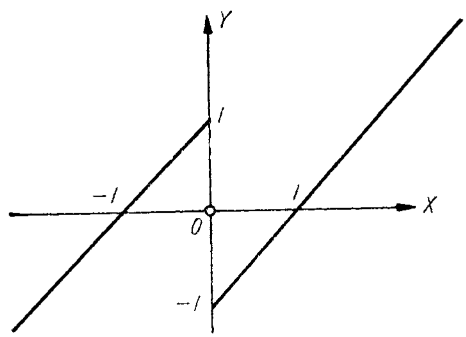

FIG. 46

of $f(x)$ for $x = c$ either sometimes does not exist, or else has nothing in common with the values of $f(x)$ for x close to c. The function $Q(t)$, for example, is not defined for $t = t_1$.

Another explanatory example may be given. We assume that a function is defined as follows in the interval $(-1, +1)$:

$$y = x + 1 \text{ for } -1 \leqslant x < 0;$$

$$y = x - 1 \text{ for } \quad 0 < x \leqslant 1; \; y = 0 \text{ for } x = 0.$$

Figure 46 shows the graph of this function; it consists of two straight sections, with their ends excluded (for $x = 0$), and a single isolated point, the origin. We now have:

$$\lim_{x\to -0} f(x) = 1; \qquad \lim_{x\to +0} f(x) = -1; \qquad f(0) = 0.$$

33. Example. We consider an example that is important later on. We take

$$y = \frac{\sin x}{x} .$$

This function is defined for all x, other than $x = 0$, for which both numerator and denominator become zero, so that the fraction loses its meaning. We shall see how y varies as x tends to zero. The magnitude of the fraction does not change when x changes sign, so that it is sufficient to find the limit of the fraction as x tends to zero through positive values, i.e. in the first quadrant. This limit exists, as we shall show. From the above remarks, the same limit is obtained for x tending to zero through negative values. We note that the theorem regarding the limit of a quotient cannot be used, since the denominator tends to zero as $x \to 0$.

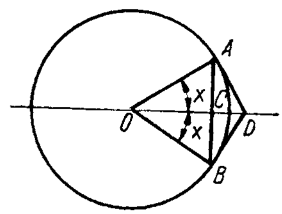

FIG. 47

We shall take x as the angle subtended at the centre of a circle of unit radius. Measuring angle in radians, we have (Fig. 47):

$$\sin x = \overline{AC}, \quad x = \frac{1}{2} \operatorname{arc} \overline{AB}, \quad \tan x = \overline{AD},$$

where $\overline{AD}$ is the tangent to the circle at the end of arc x.

Since the length of the arc is intermediate between the length of the chord and the sum of the tangents, we can write:

$$2 \sin x < 2x < 2 \tan x ,$$

whence, dividing by $2 \sin x$, we have:

$$1 < \frac{x}{\sin x} < \frac{1}{\cos x}$$

or

$$1 > \frac{\sin x}{x} > \cos x. \tag{17}$$

But as x tends to zero, $\cos x$, given by the distance $\overline{OC}$, evidently tends to unity, i.e. the variable $\sin x/x$ always lies between unity and a magnitude tending to unity, and hence [27]:

$$\lim_{x \to 0} y = \lim_{x \to 0} \frac{\sin x}{x} = 1.$$

We determine for this case the number η, encountered in condition (16).

Subtracting the three terms of (17) from unity, we have:

$$0 < 1 - \frac{\sin x}{x} < 1 - \cos x,$$

and this shows that

$$\left|1 - \frac{\sin x}{x}\right| < \varepsilon \quad \text{if} \quad |1 - \cos x| < \varepsilon.$$

Recalling that the sine of an arc in the first quadrant is less than the arc itself, we obtain:

$$1 - \cos x = 2 \sin^2 \frac{x}{2} < 2\left(\frac{x}{2}\right)^2 = \frac{x^2}{2},$$

and it is sufficient to choose:

$$\frac{x^2}{2} < \varepsilon, \quad \text{i.e.} \quad |x| < \sqrt{2\varepsilon}$$

Thus, $\sqrt{2\varepsilon}$ can act as η in the given case.

34. Continuity of functions. We have already introduced the definition of the continuity of a function at the point $x = c$, if the function is defined both at the point and in the vicinity to left and right. We give the definition again.

DEFINITION. *The function $f(x)$ is said to be continuous for $x = c$ (at the point $x = c$), if a limit of $f(x)$ exists for $x \to c$ and if this limit is equal to $f(c)$*:

$$\lim_{x \to c} f(x) = f(c) = f(\lim_{x \to c} x). \tag{18}$$

We recall that this is equivalent to the fact that *there exist limits $f(c - 0)$ and $f(c + 0)$ to left and right*, and to the fact that *these limits are equal to each other and to* $f(c)$, i.e. $f(c - 0) = f(c + 0) = f(c)$. Alternatively, the definition given above is equivalent, as we have seen [32], to: *for any given positive ε, there exists a positive η such that*

$$|f(c) - f(x)| < \varepsilon \text{ for } |c - x| < \eta. \tag{19}$$

It may be remarked that, in view of the arbitrariness of the choice of ε, we can write $|f(c) - f(x)| \leqslant \varepsilon$ in place of $|f(c) - f(x)| < \varepsilon$ in this definition. This remark applies to all previous similar definitions, and in particular, to the definition of an infinitesimal and a limit, as also to the following equivalent definition of continuity.

The difference $x - c$ is the increment of the independent variable, whilst $f(x) - f(c)$ is the corresponding increment of the function, so that the definition of continuity just given is equivalent to the

following: *a function is said to be continuous at the point $x = c$, if to an infinitesimal increment of the independent variable (from the initial value $x = c$) there corresponds an infinitesimal increment of the function.*

We note that the property of continuity, as expressed in equation (18), amounts to the possibility of finding the limit of the function by directly replacing the independent variable with its limit.

We saw from formulae (3) and (4) [28], that polynomials in x and the quotients of such polynomials, i.e. rational functions of x, are functions continuous for any x, except those for which the denominator of the rational function becomes zero.

The function $y = b$ is also obviously continuous, its value being the same for all x [12].

All the elementary functions, discussed in the first chapter (power, exponential, logarithmic, trigonometric and inverse circular), are continuous for all the x for which they exist, except those for which they tend to infinity.

For example, $\log_{10} x$ is a continuous function of x for all positive x; $\tan x$ is a continuous function of x for all x, except

$$x = (2k + 1)\frac{\pi}{2},$$

where k is any integer.

Notice further the function u^v, where u and v are continuous functions of x, u being assumed not to take negative values. This is also called an *exponential* function. It likewise has the property of continuity, except for those x for which u and v are simultaneously zero or $u = 0$ and $v < 0$.

We shall accept without proof what has been said about the continuity of the elementary functions, although proof is of course required, and can in fact be given with complete rigour. We shall later examine the question in detail.

It can easily be shown that *the sum or product of any finite number of continuous functions is itself a continuous function; the same is true of the quotient of two continuous functions except for those values of the independent variable for which the denominator tends to zero.*

We only consider the case of a quotient. We assume that functions $\varphi(x)$ and $\psi(x)$ are continuous for $x = a$ and that $\psi(a) \neq 0$. We take the function

$$f(x) = \frac{\varphi(x)}{\psi(x)}.$$

Using the theorem concerning the limit of a quotient, we obtain:

$$\lim_{x\to a} f(x) = \frac{\lim\limits_{x\to a}\varphi(x)}{\lim\limits_{x\to a}\psi(x)} = \frac{\varphi(a)}{\psi(a)} = f(a),$$

which proves the continuity of the quotient $f(x)$ for $x = a$.

We note one simple example. If $y = \sin x$ is a continuous function of x, $y = b \sin x$, where b is a constant, will also be continuous, being the product of the continuous functions $y = b$ (see above) and $y = \sin x$.

We turn again now to the function $y = \sin x/x$. This is not defined for $x = 0$, but we know that $\lim\limits_{x\to 0} y = 1$. Hence, if we put $y = 1$ for $x = 0$, y will be a continuous function at the point $x = 0$.

Such a process of finding the limit of a function for x tending to its point of indeterminacy is called *disclosing the indeterminacy*, and the limit itself, if it exists, is sometimes called a *true value* of the function at this point of indeterminacy. We shall have many examples later on of the disclosure of indeterminacies.

35. The properties of continuous functions. We defined above the continuity of a function for a given value of x. We now suppose that the function is defined in a finite interval $a \leqslant x \leqslant b$. *If it is continuous for any given x in this interval, we say that it is continuous in the interval* (a, b). We note here that continuity of the function at the ends of the interval, $x = a$ and $x = b$, consists in:

$$\lim_{x\to a+0} f(x) = f(a), \quad \lim_{x\to b-0} f(x) = f(b).$$

All continuous functions have the following properties:

1. *If the function $f(x)$ is continuous in the interval (a, b), there exists at least one value of x in this interval at which $f(x)$ takes its maximum value, and at least one value of x for which the function takes its minimum value.*

2. *If the function $f(x)$ is continuous in the interval (a, b), with $f(a) = m$ and $f(b) = n$, and if k is any number lying between m and n, there exists at least one x in the interval such that $f(x) = k$; and in particular, if $f(a)$ and $f(b)$ have opposite signs, there exists at least one x in the interval such that $f(x)$ is zero.*

These two properties are immediately clear, if we note that the graph corresponding to a continuous function is a continuous curve. This remark cannot serve as a proof, of course. The concept itself of

a continuous curve, obvious at first sight, is seen to be unusually complex on closer inspection. The rigorous proof of the two properties mentioned, as also of the third, to follow, is based on the theory of irrational numbers. We accept these properties without proof.

In subsequent paragraphs of the present section, we study the basis of the theory of irrational numbers and the relationship of this theory to the theory of limits and to the properties of continuous functions.

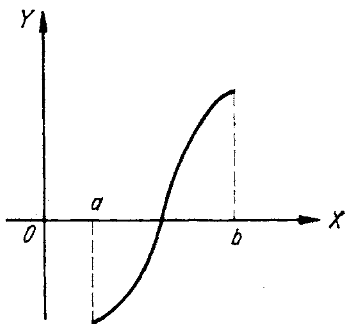

FIG. 48

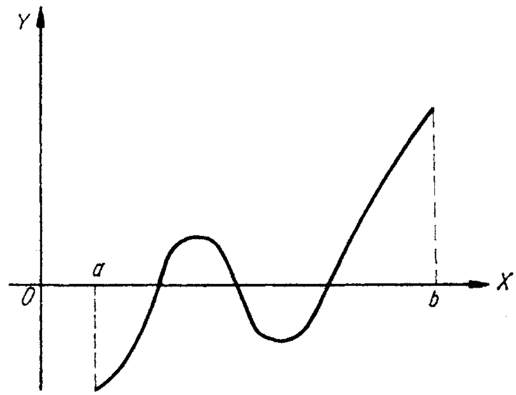

FIG. 49

We may remark that the second property of continuous functions can also be formulated thus: on continuous variation of x from a to b, the continuous function $f(x)$ passes at least once through every number lying between $f(a)$ and $f(b)$.

Figures 48 and 49 show the graphs of functions, continuous in the interval (a, b), for which $f(a) < 0$ and $f(b) > 0$. In Fig. 48 the graph cuts the axis OX once, and $f(x)$ is zero for the corresponding x. There are three such values of x, instead of one, in the case of Fig. 49.

We now pass to the third property of continuous functions, which is less obvious than the two previous ones.

3. If $f(x)$ is continuous in the interval (a, b), and if $x = x_0$ is a certain value of x in this interval, by condition (19) [34] (replacing c by x_0), for any given positive ε there exists an η, of course depending on ε, such that

$$|f(x) - f(x_0)| < \varepsilon, \text{ if } |x - x_0| < \eta,$$

it naturally being assumed that x also lies in this interval. (If, for example, $x_0 = a$, x must be greater than a, and if $x_0 = b$, $x < b$.) But the number η can depend, not only on ε, but also on just what

value of $x = x_0$ we take in the interval. The third property of continuous functions consists in the fact that, for any given ε, there exists the same η for all x_0 in the interval (a, b). In other words, *if $f(x)$ is continuous in the interval (a, b), for any given positive ε there exists a positive η such that*

$$|f(x'') - f(x')| < \varepsilon \tag{20}$$

for any two values x'' and x' in the interval (a, b) which satisfy the inequality

$$|x'' - x'| < \eta. \tag{21}$$

This property is referred to as *uniform continuity*. Thus, *if a function is continuous in an interval (a, b), it is uniformly continuous in this interval.*

We again remark, that we assume $f(x)$ to be continuous, not only for all x inside the interval (a, b), but also for $x = a$ and $x = b$.

We shall further illustrate the property of uniform continuity by a simple example. We first rewrite the above inequality in another form, replacing the symbol x' by x, and x'' by $(x + h)$. Now $x'' - x' = h$ is the increment of the independent variable, and $f(x + h) - f(x)$ is the corresponding increment of the function. The property of uniform continuity now becomes:

$$|f(x + h) - f(x)| < \varepsilon \quad \text{if} \quad |h| < \eta,$$

where x and $(x + h)$ are any two points in the interval (a, b).

Take the example of the function:

$$f(x) = x^2.$$

We now have:

$$f(x + h) - f(x) = (x + h)^2 - x^2 = 2xh + h^2.$$

For any given x, the expression $(2xh + h^2)$ for the increment of our function obviously tends to zero, as the increment of the independent variable tends to zero. This is a further confirmation (cf. [34]) that the function in question is continuous for every x. It will be continuous, for instance, in the interval $-1 < x < 2$. We show that it is uniformly continuous in this interval. We have to satisfy the inequality:

$$|2xh + h^2| < \varepsilon \tag{22}$$

for suitable choice of η in the inequality $|h| < \eta$, where x and $(x + h)$ must lie in the interval $(-1, 2)$. We have:

$$|2xh + h^2| < |2xh| + h^2 = 2|x||h| + h^2.$$

The maximum value of $|x|$ in the interval is two, and hence we can rewrite the above inequality with greater force as:

$$|2xh + h^2| \leqslant 4|h| + h^2.$$

We shall always take $|h| < 1$. Then $h^2 < |h|$, and we can write the above in the form:

$$|2xh + h^2| < 4|h| + |h|$$

or

$$|2xh + h^2| < 5|h|.$$

The inequality (22) will certainly be satisfied, if we take $|h|$ on the condition $5|h| < \varepsilon$. Thus, h must satisfy two inequalities:

$$|h| < 1 \text{ and } |h| < \frac{\varepsilon}{5}.$$

We can thus take for η the least of the two numbers 1 and $\varepsilon/5$. For small ε (in fact, $\varepsilon < 5$), we must take $\eta = \varepsilon/5$, and this η is evidently the same, for a given ε, for all x in the interval $(-1, 2)$.

The property mentioned cannot obtain in the case of discontinuous functions, or those continuous only inside an interval. Take the function, the graph of which is shown in Fig. 46. It is defined in the interval $(-1, +1)$ and has a discontinuity at $x = 0$. It has values as close as desired to unity, but it does not take the value unity, or values greater than unity. There is thus no maximum among the values of this function. Similarly, there is no minimum. The elementary function $y = x$ does not take either a maximum or minimum value inside the interval (0.1). If it is considered in the closed interval (0,1), it reaches its minimum value at $x = 0$, and its maximum at $x = 1$. Take another function, $f(x) = \sin(1/x)$ continuous in the interval $0 < x \leqslant 1$, open on the left. As x tends to zero, the argument $1/x$ increases indefinitely, and $\sin(1/x)$ oscillates between (-1) and $(+1)$, having no limit as $x \to +0$. We show that this function is not uniformly continuous in the interval $0 < x \leqslant 1$. We take two values: $x' = 1/n\pi$ and $x'' = 2/(4n + 1)\pi$, where n is a positive integer. Both values lie in the interval for any choice of n. Further, we have:

$$f(x') = \sin n\pi = 0;$$

$$f(x'') = \sin\left(2n\pi + \frac{1}{2}\pi\right) = 1.$$

Thus:

$$f(x'') - f(x') = 1$$

and

$$x'' - x' = \frac{2}{(4n + 1)\pi} - \frac{1}{n\pi}.$$

As the positive integer n tends to infinity, the difference $x'' - x'$ tends to zero, whilst $f(x'') - f(x')$ remains equal to 1. It is thus evident that there does not exist a positive η, such that, in the interval $0 < x \leqslant 1$, (21) implies $|f(x'') - f(x')| < 1$; this corresponds to choosing $\varepsilon = 1$ in formula (20).

Take the function $f(x) = x \sin(1/x)$. The first term of the product tends to zero as $x \to +0$, whilst the absolute value of the second, $\sin(1/x)$, does not exceed unity; hence [32], $f(x) \to 0$ as $x \to +0$. The second term has no meaning for $x = 0$, but if we complete the definition of our function by taking $f(0) = 0$, i.e. if we take $f(x) = x \sin(1/x)$ for $0 < x \leqslant 1$ and $f(0) = 0$, we obtain a function continuous in the closed interval $(0,1)$. The functions $\sin(1/x)$ and $x \sin(1/x)$ are evidently continuous for any x, excepting zero.

36. Comparison of infinitesimals and of infinitely large magnitudes. If α and β are two magnitudes, simultaneously tending to zero, the theorem regarding the limit of a quotient cannot be used for finding the limit of the ratio β/α. We shall assume that the variables α and β, whilst tending to zero, do not take the value zero. If the ratio β/α tends to a finite limit, differing from zero, the ratio α/β will also tend to a finite limit, differing from zero. We say in this case that β and α are *infinitesimals of the same order*. If the ratio β/α has a limit at zero, we say that β is an infinitesimal of higher order in comparison with α, or that α is an infinitesimal of lower order in comparison with β. If the ratio β/α tends to infinity, α/β tends to zero, i.e. β is of lower order compared with α, and α of higher order compared with β. It is easy to show that, *if α and β are infinitesimals of the same order, and γ is an infinitesimal of higher order compared with α, γ is also of higher order as regards β*. By hypothesis $\gamma/\alpha \to 0$, and α/β has a finite limit, differing from zero. From the self evident equation $\gamma/\beta = \gamma/\alpha \cdot \alpha/\beta$, and using the theorem regarding the limit of a product, it follows at once that $\gamma/\beta \to 0$, which proves our statement.

We note an important particular case of infinitesimals of the same order. If $\alpha/\beta \to 1$ (so that also $\beta/\alpha \to 1$), infinitesimals α and β are referred to as equivalent. It follows at once from the equation

$$\frac{\beta - \alpha}{\alpha} = \frac{\beta}{\alpha} - 1,$$

that *the equivalence of α and β implies that the difference $\beta - \alpha$ is an infinitesimal of higher order than α*. It similarly follows from the equation

$$\frac{\beta - \alpha}{\beta} = 1 - \frac{\alpha}{\beta}$$

that their equivalence implies that $\beta - \alpha$ is an infinitesimal of higher order than β.

If β/a^k, where k is a positive constant, tends to a finite limit, differing from zero, we say that β is an infinitesimal of order k with respect to a. If $\beta/a \to c$, where c is a number, not zero, $|\beta/c\,a^k| \to 1$, i.e. β and ca^k are equivalent infinitesimals, and therefore, $\gamma = \beta - ca^k$ is an infinitesimal of higher order than β (or than ca^k). If a is taken as the basic infinitesimal, the equation $\beta = ca^k + \gamma$, where γ is an infinitesimal of higher order than ca^k, represents the isolation from the infinitesimal β of the infinitesimal term ca^k (of the simplest form with respect to a), in such a way that the remainder is an infinitesimal γ of higher order than β (or than ca^k).

An analogous comparison can be made of the infinitely large magnitudes u and v. If v/u tends to a limit, finite and not zero, we say that u and v are infinitely large magnitudes of the same order. If $v/u \to 0$, then $u/v \to \infty$. We say in this case that v is of a lower order of greatness with respect to u, or that u is of a higher order of greatness with respect to v. If $v/u \to 1$, the infinitely large magnitudes are said to be equivalent. If v/u^k, where k is a positive constant, has a limit, which is finite and not zero, we say that v is of the kth order of greatness with respect to u. All the above remarks about infinitesimals apply for infinitely large magnitudes.

We further remark, that if the ratio β/a or v/u has no limit at all, the corresponding infinitesimals or large order magnitudes are said to be incomparable.

37. Examples.

1. We saw above that

$$\lim_{x \to 0} \frac{\sin x}{x} = 1,$$

i.e. $\sin x$ and x are equivalent infinitesimals, and therefore $\sin x - x$ is an infinitesimal of higher order than x. We see later, that this difference is equivalent to $-x^3/6$, i.e. it is an infinitesimal of the third order with respect to x.

2. We show that the difference $1 - \cos x$ is an infinitesimal of the second order with respect to x. We have in fact, on using a well-known trigonometric formula and with a simple rearrangement,

$$\frac{1-\cos x}{x^2} = \frac{2\sin^2 \dfrac{1}{2}x}{x^2} = \frac{1}{2}\left(\frac{\sin \dfrac{1}{2}x}{\dfrac{1}{2}x}\right)^2 .$$

If $x \to 0$, $\alpha = x/2$ also tends to 0, and as we have shown:

$$\lim_{x\to 0} \frac{\sin \frac{1}{2} x}{\frac{1}{2} x} = \lim_{\alpha \to 0} \frac{\sin \alpha}{\alpha} = 1,$$

and hence,

$$\lim_{x\to 0} \frac{1-\cos x}{x^2} = \frac{1}{2},$$

i.e. in fact, $1 - \cos x$ is an infinitesimal of the second order with respect to x.

3. From the expression

$$\sqrt{1+x} - 1 = \frac{x}{\sqrt{1+x}+1}$$

we have:

$$\frac{\sqrt{1+x}-1}{x} = \frac{1}{\sqrt{1+x}+1},$$

whence

$$\lim_{x\to 0} \frac{\sqrt{1+x}-1}{x} = \frac{1}{2},$$

i.e. $\sqrt{1+x} - 1$ and x are infinitesimals of the same order, $\sqrt{1+x} - 1$ being equivalent to $x/2$.

4. We show that a polynomial of degree $m > 1$ is an infinitely large magnitude of order m with respect to x. In fact,

$$\lim_{x\to\infty} \frac{a_0 x^m + a_1 x^{m-1} + \cdots + a_{m-1} x + a_m}{x^m} =$$

$$= \lim_{x\to\infty} \left(a_0 + \frac{a_1}{x} + \cdots + \frac{a_{m-1}}{x^{m-1}} + \frac{a_m}{x^m}\right) = a_0.$$

It can easily be seen that two polynomials of the same degree are infinitely large magnitudes of the same order, for $x \to \infty$. The limit of their ratio is the ratio of the coefficients of their highest terms. For example:

$$\lim_{x\to\infty} \frac{5x^2 + x - 3}{7x^2 + 2x + 4} = \lim_{x\to\infty} \frac{5 + \frac{1}{x} - \frac{3}{x^2}}{7 + \frac{2}{x} + \frac{4}{x^2}} = \frac{5}{7}.$$

If the two polynomials are of different degree, the one of higher degree is an infinitely large magnitude of higher order with respect to the other, for $x \to \infty$.

38. The number e. Our present example is important later on: we consider the variable taking the values

$$\left(1 + \frac{1}{n}\right)^n,$$

where n tends to $+\infty$, increasing through positive integers. Using Newton's binomial formula, we obtain:

$$\left(1+\frac{1}{n}\right)^n = 1+\frac{n}{1}\times\frac{1}{n}+\frac{n(n-1)}{2!}\times\frac{1}{n^2}+$$

$$+\frac{n(n-1)(n-2)}{3!}\times\frac{1}{n^3}+$$

$$+\dots+\frac{n(n-1)(n-2)\dots(n-k+1)}{k!}\times\frac{1}{n^k}+\dots+$$

$$+\frac{n(n-1)(n-2)\dots2\cdot1}{n!}\times\frac{1}{n^n}=$$

$$=1+1+\frac{1}{2!}\left(1-\frac{1}{n}\right)+\frac{1}{3!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)+\dots+$$

$$+\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right)+\dots+$$

$$+\frac{1}{n!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{n-1}{n}\right).$$

The sum written above contains $(n+1)$ positive terms. As the integer n increases, the number of terms increases and each term itself also increases, since in the expression for the general term:

$$\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right)\dagger$$

$k!$ remains unchanged, whilst the differences in brackets increase with increasing n. We thus see that the variable in question increases with increasing n; so that it is sufficient to show that the variable is bounded, in order to prove that its limit exists.

We replace all the differences appearing in the general term by unity, and all the factors of $k!$, starting with 3, by 2. The general

† The product $\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\dots\left(1-\frac{k-1}{n}\right)$ is obtained from the fraction $\frac{n(n-1)(n-2)\dots(n-k+1)}{n^k}$ if, noting that there are k terms n in the denominator, each of the k terms of the product on top is divided by n.

term is evidently now increased, and we shall have, on using the formula for the sum of the terms of a geometrical progression:

$$\left(1+\frac{1}{n}\right)^n < 1+1+\frac{1}{2}+\frac{1}{2^2}+\ldots+\frac{1}{2^{k-1}}\ldots+\frac{1}{2^{n-1}} =$$

$$= 1 + \frac{1-\frac{1}{2^n}}{1-\frac{1}{2}} = 3 - \frac{1}{2^{n-1}} < 3,$$

i.e. the variable $(1+1/n)^n$ is bounded. We denote its limit by the letter e:

$$\lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^n = \mathrm{e} \quad (n \text{ is a positive integer}). \tag{23}$$

This limit is evidently not greater than 3.

We now show that the expression $(1+1/x)^x$ tends to the same limit e, if x tends to $+\infty$, taking any values.

Let n be the greatest integer included in x, i.e.

$$n \leqslant x < n+1.$$

The number n evidently tends to $+\infty$ along with x. On noting that a power term increases, both with increase of the positive base, greater than unity, and with increase of the exponent of the power, we can write:

$$\left(1+\frac{1}{n+1}\right)^n < \left(1+\frac{1}{x}\right)^x < \left(1+\frac{1}{n}\right)^{n+1} \tag{24}$$

But by equation (23):

$$\lim_{n\to+\infty}\left(1+\frac{1}{n+1}\right)^n = \lim_{n\to+\infty}\frac{\left(1+\frac{1}{n+1}\right)^{n+1}}{\left(1+\frac{1}{n+1}\right)} = \frac{\mathrm{e}}{1} = \mathrm{e}$$

and

$$\lim_{n\to+\infty}\left(1+\frac{1}{n}\right)^{n+1} = \lim_{n\to+\infty}\left[\left(1+\frac{1}{n}\right)^n\left(1+\frac{1}{n}\right)\right] = \mathrm{e}.$$

Thus, the extreme terms of inequality (24) tend to the limit e, and hence the middle term must tend to the same limit, i.e.

$$\lim_{x\to+\infty}\left(1+\frac{1}{x}\right)^x = \mathrm{e}. \tag{25}$$

We now consider the case when x tends to $-\infty$.

We introduce a new variable y in place of x, putting

$$x = -1 - y\,,$$

whence

$$y = -1 - x.$$

It is evident from the last equation that y tends to $+\infty$ as x tends to $-\infty$.

On changing the variables in the expression $(1 + 1/x)^x$ and noting equation (25), we obtain:

$$\lim_{x\to-\infty}\left(1 + \frac{1}{x}\right)^x = \lim_{y\to+\infty}\left(\frac{-y}{-1-y}\right)^{-1-y} =$$

$$= \lim_{y\to+\infty}\left(\frac{1+y}{y}\right)^{1+y} = \lim_{y\to+\infty}\left[\left(1 + \frac{1}{y}\right)^y\left(1 + \frac{1}{y}\right)\right] = \mathrm{e}\cdot 1 = \mathrm{e}.$$

If x tends to ∞, with either sign, i.e. $|x| \to +\infty$, it follows from the above that here also:

$$\lim_{x\to\infty}\left(1 + \frac{1}{x}\right)^x = \mathrm{e}. \tag{26}$$

We shall later give a suitable method for calculating e to any degree of accuracy. Clearly, it is an irrational number; we have, to an accuracy of seven decimal places: $\mathrm{e} = 2.7182818\ldots$

We can now easily find the limit of $(1 + k/x)^x$, where k is a given number. Using the continuity of a power function, we obtain:

$$\lim_{x\to\infty}\left(1 + \frac{k}{x}\right)^x = \lim_{x\to\infty}\left[\left(1 + \frac{1}{x/k}\right)^{x/k}\right]^k = \lim_{x\to\infty}\left[\left(1 + \frac{1}{y}\right)^y\right]^k = \mathrm{e}^k$$

where y denotes x/k, and tends to infinity along with x.

An expression of the form $(1 + k/n)^n$ is encountered in compound interest theory.

We suppose that an increment of capital occurs annually. If capital a returns an interest annually of p per cent, the accumulated capital in the course of a year will be:

$$a(1 + k)\,,$$

where

$$k = \frac{p}{100}\,;$$

after another year has elapsed, it will be:

$$a(1 + k)^2;$$

and in general, after the lapse of m years, it will be:

$$a(1 + k)^m.$$

We now suppose that the increment of capital takes place every $1/n$ of a year. The number k is now diminished n times, since the percentage interest is counted over a year, whilst the number of intervals of time is increased n times; so that the accumulated capital over m years will be:

$$a\left(1 + \frac{k}{n}\right)^{mn}.$$

Finally, let n tend to infinity, i.e. an increment of capital occurs in every smallest possible interval of time, and in the limit, continuously. After the lapse of m years, the accumulated capital will be:

$$\lim_{n\to\infty} a\left(1 + \frac{k}{n}\right)^{mn} = \lim_{n\to\infty} a\left[\left(1 + \frac{k}{n}\right)^n\right]^m = a\mathrm{e}^{km}.$$

The number e is used as a base of logarithms. These are referred to as *natural logarithms* and are here denoted by the simple sign log without indicating the base.

For x tending to zero, both numerator and denominator in the expression $\log(1 + x)/x$ tend to zero. Let us examine this indeterminate form. We introduce a new variable y, putting

$$x = \frac{1}{y}, \quad \text{i.e.} \quad y = \frac{1}{x},$$

whence evidently, as $x \to 0$, y tends to infinity. Substituting the new variable, and making use of the continuity of a logarithm and formula (26), we obtain:

$$\lim_{x\to 0} \frac{\log(1 + x)}{x} = \lim_{y\to\infty} y \log\left(1 + \frac{1}{y}\right) = \lim_{y\to\infty} \log\left(1 + \frac{1}{y}\right)^y = \log \mathrm{e} = 1.$$

The advantage of the present choice of a base of logarithms is clear from this. Just as, using radian measure of angles, the true value of $(\sin x)/x$ is unity for $x = 0$, *in the case of natural logarithms the true value of* $\log(1 + x)/x$ *is also unity for* $x = 0$.

The following relationship follows from the definition of logarithm:

$$N = a^{\log_a N}.$$

Taking logarithms to base e in this equation, we obtain:

$$\log N = \log_a N \cdot \log a \;\text{ or }\; \log_a N = \log N \cdot \frac{1}{\log a}.$$

This relationship gives the logarithm of a number N to any base a in terms of its natural logarithm. The factor $M = 1/\log a$ is called the *modulus* of the system of logarithms to base a, and for $a = 10$ it is given with an accuracy of seven decimal places by:

$$M = 0.4342945 \ldots$$

39. Unproved hypotheses. We left several hypotheses unproved, in dealing with the theory of limits, and we now give them again: the existence of a limit for a monotonic bounded variable [30], the necessary and sufficient condition for the existence of a limit (Cauchy's test) [31] and the three properties of a function continuous in a closed interval [35]. The proofs of these hypotheses are based on the theory of real numbers and the operations on them. The following paragraphs will deal with this theory, and with the proofs of the above hypotheses.

We introduce another new concept, and formulate a further hypothesis, the proof of which will also be given below. If we have a set, consisting of a finite number of real numbers (for example, we might have a thousand real numbers), there will be both a maximum and a minimum amongst them. On the other hand, if we have an infinite set of real numbers, such that, furthermore, all the numbers belong to a definite interval, there will not always be a maximum and minimum amongst them. For example, if we take the set of all real numbers, lying between 0 and 1, but at the same time exclude from this set the numbers 0 and 1 themselves, there will be neither a maximum nor a minimum in the set. Whatever number we take, near to but less than, unity, we can always find a second number, lying between this first number and unity. The numbers 0 and 1 in this case, whilst not belonging to our set of numbers, have the following property in relation to it: there is no number greater than unity among the set, but for any given positive number ε there exists a number greater than $(1 - \varepsilon)$. Similarly, there is no number less than zero among the set, but for any given positive number ε there exists a number less than $(0 + \varepsilon)$. The numbers 0 and 1 are called the *strict lower* and *strict upper bounds* of the set of real numbers in question.

We pass from this example to the general case. Let E be some set of real numbers. We say that it is *bounded above*, if there exists a number M, such that all numbers belonging to E are not greater than M. Similarly, we say that the set is *bounded below*, if there exists a number m, such that all numbers belonging to E are not

less than m. If a set is bounded above and below, we simply say that it is *bounded.*

DEFINITION. *The strict upper bound of a set E is defined as the number β (if it exists), such that no member of E is greater than β, whereas, for any given positive ε, there is a member greater than $(\beta - \varepsilon)$. The strict lower bound of the set is defined as the number α (if it exists), such that no member of E is less than α, whereas for any given positive ε there is a member less than $(\alpha + \varepsilon)$.*

If the set E is not bounded above, i.e. if there exists a member of E greater than any given number, the set cannot have a strict upper bound. Similarly, if E is not bounded below, it cannot have a strict lower bound. If there exists a maximum among the numbers of the set, this is evidently the strict upper bound of the set. Similarly, if there is a minimum among the numbers of the set, this is the strict lower bound of set E. But as we have seen, there is not always a maximum or minimum among the members of an infinite set. It can be shown, however, that *there is always a strict upper bound for a set bounded above, and always a strict lower bound for a set bounded below.* We also note a direct consequence of the definition of strict bounds, that the strict upper and strict lower bounds must be unique.

Later on, we shall often make use of the hypotheses indicated in the present paragraph. The next paragraph, in small print, can be omitted at a first reading.

40. Real numbers. We begin by dealing with the theory of real numbers. We start out from the set of all rational numbers, integral and fractional, positive and negative. All these rational numbers can be arranged in increasing order. If this is done, and a and b are any two distinct rational numbers, any desired number of rational numbers can be found between them. Let $a < b$, and let us introduce the positive rational number $r = (b - a)/n$, where n is any positive integer. The rational numbers $a + r, a + 2r, a + 3r, \ldots a + (n - 1)r$ lie between a and b, and in view of the arbitrariness of choice of the integer n, our statement is proved.

We define a *section in the domain of real numbers* as any division of all rational numbers into two classes, such that any number of one (the first) class is less than any number of the other (second) class. Evidently, in this case, if a given number belongs to the first class, every number less than it also belongs to the first class, and if a given number belongs to the second class, every number greater than it also belongs to the second class.

We suppose that there is a greatest among the numbers of the first class. In this case, by the property mentioned of the set of rational numbers, it can be asserted that there is no minimum among the numbers of the second class. Similarly, if there is a least among the numbers of the second class,

there is no greatest among the numbers of the first class. We call the section a *section of the first kind*, if there is a greatest among the numbers of the first class, or if there is a least among numbers of the second class. It is easy to construct such a section. We take any rational number b and place all rational numbers less than b in the first class, and all rational numbers greater than b in the second class, whilst b itself is placed either in the first class (where it will be the greatest) or in the second class (where it will be the least). Taking every possible rational number as b, we obtain every possible section of the first kind. We shall say that such a section of the first kind defines the rational number b, this being the greatest in the first, or the least in the second, class.

But there exists a *section of the second kind*, where there is no greatest in the first, or least in the second, class. We construct one such section, as an example. We put in the first class all negative rational numbers, zero, and those positive rational numbers whose square is less than two, whilst we put in the second class all those positive rational numbers whose square is greater than two. Since there is no rational number whose square is equal to two, all the rational numbers are now assigned, and we have a certain section. We shall show that there is no greatest number in the first class. To do this, it is sufficient to show that, if a belongs to the first class, there exists a number greater than a, also belonging to the first class. This is evident if a is negative or zero; so we suppose $a > 0$. Since it belongs to the first class, $a^2 < 2$. We bring in a positive rational number $r = 2 - a^2$, and we show that a positive rational number x can be defined, small enough for $(a + x)$ also to belong to the first class, i.e. for us to have the inequality:

$$2 - (a + x)^2 > 0 \quad \text{or} \quad r - 2ax - x^2 > 0\,,$$

i.e. it amounts to us finding a positive rational number which satisfies the inequality:

$$x^2 + 2ax < r.$$

Taking $x < 1$, we have $x^2 < x$, and hence, $x^2 + 2ax < x + 2ax = (2a + 1)\, x$ i.e. it is sufficient for us to satisfy the inequality:

$$(2a + 1)\, x < r\,,$$

so that x is defined by the two inequalities:

$$x < 1 \text{ and } x < \frac{r}{2a + 1}\,.$$

We can evidently find as many positive rational numbers x as desired, satisfying both these inequalities. It can be shown in precisely the same way, that there is no least number in the second class of our section. We have thus constructed an example of a section of the second kind. The following hypothesis is a turning-point in the theory: we suppose that *every section of the second kind defines a certain new entity, i.e. an irrational number*. Different sections of the second kind define different irrational numbers. In the above example of a section of the second kind, it is easy to surmise that the irrational number defined is that usually denoted by $\sqrt{2}$.

We can now arrange all the irrational numbers thus introduced, together with the previous rational numbers, in increasing order, as is done intuitively with the points of the directed axis OX. If a is a certain irrational number, we denote the first and second classes of the section which defines it by $\mathrm{I}(a)$ and $\mathrm{II}(a)$ respectively. We reckon a as greater than any number of $\mathrm{I}(a)$, and less than any number of $\mathrm{II}(a)$. Any irrational can thus be compared with any rational number. It remains to define the concepts of greater than and less than for any two distinct irrational numbers a and β. Since a and β are distinct, classes $\mathrm{I}(a)$ and $\mathrm{I}(\beta)$ do not coincide, and one class is contained in the other. We suppose that $\mathrm{I}(a)$ is contained in $\mathrm{I}(\beta)$, i.e. every number of $\mathrm{I}(a)$ belongs to $\mathrm{I}(\beta)$, whilst there is a number of $\mathrm{I}(\beta)$ belonging to $\mathrm{II}(a)$. We take it that here, by definition, $a < \beta$. In this way, the set of all rational and irrational numbers, or in other words, the *set of all real numbers* is arranged in order. Using the definitions given above, it is now easy to show that, if a, b, c are real numbers, and $a < b$, $b < c$, then $a < c$.

We note above all one elementary consequence of the definitions given. Let a be a certain set number. Since there is no greatest number in class $\mathrm{I}(a)$, and no least number in class $\mathrm{II}(a)$, it is immediately evident that any desired number of rational numbers may be set between a and any given rational number a. Now let $a < \beta$ be two distinct irrational numbers. Part of the rational numbers of $\mathrm{I}(\beta)$ enter into $\mathrm{II}(a)$, and hence it immediately follows that it is also possible to place any desired number of rational numbers between a and β, i.e. in general, *any desired number of rational numbers can be placed between two distinct real numbers.*

We now pass to the proof of the basic theorem of the theory of irrational numbers. We take the aggregate of all real numbers and make some section of it, i.e. we assign all real numbers, both rational and irrational, to two classes I and II, such that any number of I is less than any number of II. We show that there must now be either a greatest number in class I, or a least number in class II (one excludes the other, as above, for a section of the domain of rational numbers). For this, we denote by I′ the set of all rational numbers of I, and by II′, the set of all rational numbers of II. Classes I′ and II′ define a certain section of the domain of rational numbers, and this section defines a real number a (rational or irrational). We suppose for clarity that this a lies in class I, with the above assignment of all real numbers into two classes. We show that a must be the greatest number of class I. If this were untrue, there would exist a real number β of class I, greater than a. We take some rational number r, lying between a and β, i.e. $a < r < \beta$. It must belong to class I, and hence, to class I′.

Thus, the number r, greater than a, is in the first class of the section (I′, II′) defining a. But this is impossible; and hence our assertion that a is the greatest number of class I must be true. It can similarly be shown, that if a belongs to class II, it must be the least number there.

We have thus proved the following basic theorem:

FUNDAMENTAL THEOREM. *For any section, made in the domain of real numbers, either the first class must contain a greatest number, or the second class a least number.*

A simple geometrical meaning can easily be attached to all the arguments of the present paragraph. To start with, we take on the axis OX only the points with rational abscissae. A section in the domain of rational numbers corresponds to cutting OX into two semi-axes. If the cutting is at a point with a rational abscissa, the section obtained is of the first kind, the abscissa of the point of cutting being itself counted as either in the first, or in the second, class. If the cutting occurs at a point for which there is no corresponding rational abscissa, a section of the second kind is obtained, defining an irrational number which is taken to be the abscissa of the point of cutting. After filling in such empty points with irrational abscissae, every division of the axis now occurs at a point with a certain real abscissa. All this is mere geometrical illustration, and lacks the force of a proof. Using the definition given of an irrational number α, it is easy to form the infinite decimal fraction corresponding to this number [2]. Every finite piece of this fraction must belong to $\mathrm{I}(\alpha)$, but if the last figure of this piece is increased by unity, the resultant rational number must belong to $\mathrm{II}(\alpha)$.

41. The operations on real numbers. The theory of irrational numbers contains, in addition to the definitions and the basic theorem given above, definitions of the operations on irrational numbers, and a study of the properties of these operations. We shall define the operations by making use of sections in the domain of rational numbers; moreover, since these sections define not only irrational, but also rational numbers (sections of the first kind), the definitions of the operations will suit all real numbers in general, and will coincide with what is known in the case of rational numbers. Our discussion is limited to general remarks in the present section.

We make a preliminary lemma. Let α be a certain real number. We take some (small) positive rational number r, then a rational number a of $\mathrm{I}(\alpha)$, and form the arithmetic progression:

$$a, a + r, a + 2r, \ldots, a + nr, \ldots$$

For n large, $(a + nr)$ belongs to $\mathrm{II}(\alpha)$; hence, there will exist a positive integer k, such that $[a + (k - 1) r]$ lies in $\mathrm{I}(\alpha)$, and $(a + kr)$ lies in $\mathrm{II}(\alpha)$, i.e.:

LEMMA. *For any given section of the rational numbers, there exist numbers in the two classes differing by any given positive rational number r, however small.*

We now pass to defining addition. Let α and β be two real numbers. Let a be any number of $\mathrm{I}(\alpha)$, a' be any number of $\mathrm{II}(\alpha)$, b any number of $\mathrm{I}(\beta)$, b' any number of $\mathrm{II}(\beta)$. We form all the possible sums $(a + b)$ and $(a' + b')$. In every case we have: $a + b < a' + b'$. We carry out a new section of the rational numbers, putting all rationals greater than all $(a + b)$ in the second class, and all remaining rationals in the first class. Any number of the first class is now less than any number of the second class, all the numbers $(a + b)$ falling in the first class, and all $(a' + b')$ in the second class. This new section defines a certain real number, which we call the sum $(\alpha + \beta)$. This number is evidently greater than or equal to all $(a + b)$, and less than or equal to all $(a' + b')$. Noting that, by the above lemma, numbers a and a', as also

b and b', can differ from each other by any given small positive rational number, it is easy to see that there can exist only one number, satisfying the above-mentioned inequality. We can at once verify that addition satisfies the ordinary laws, known for rational numbers:

$$\alpha + \beta = \beta + \alpha; \ (\alpha + \beta) + \gamma = \alpha + (\beta + \gamma); \ \alpha + 0 = \alpha .$$

For example, to obtain $(\beta + \alpha)$, we should have to construct the sums $(b + a)$ and $(b' + a')$ instead of $(a + b)$ and $(a' + b')$, in which case the former sums coincide with the latter, since addition of rational numbers is known to obey the rule of transposition.

Let α be a certain real number. We define the number $(-\alpha)$ by the section carried out as follows: we put in the first class all the rational numbers of $\mathrm{II}(\alpha)$ with changed sign, and in the second class, all numbers of $\mathrm{I}(\alpha)$ with changed sign. This is in fact a section in the domain of rational numbers; and we have for $(-\alpha)$, as is easily verified:

$$-(-\alpha) = \alpha; \ \alpha + (-\alpha) = 0 .$$

Clearly, if $\alpha < 0$, $(-\alpha) > 0$, and conversely. Provided $\alpha \neq 0$, we say that whichever of the numbers α and $(-\alpha)$ is greater than zero is its absolute value. As before, $|\alpha|$ denotes the absolute value of α.

We now pass to multiplication. Let α and β be two positive real numbers, i.e. $\alpha > 0$ and $\beta > 0$. Let a be any *positive number* of $\mathrm{I}(\alpha)$, b be any *positive number* of $\mathrm{I}(\beta)$, and a', b' any numbers of $\mathrm{II}(\alpha)$ and $\mathrm{II}(\beta)$ (they must also be positive). We construct a new section, putting all rational numbers greater than all products ab in the second class, and all the remaining rational numbers in the first class. All ab fall in the first class, and all $a'b'$ in the second class. The new section defines a certain real number, which we call the product $\alpha\beta$. This number is greater than or equal to all ab and not greater than all $a'b'$, and there is only one real number satisfying these inequalities.

In the case of negative α and/or β, we carry out multiplication as in the previous case, whilst introducing into the definition of multiplication the usual rule of signs, i.e. we put $\alpha\beta = \pm |\alpha||\beta|$, taking the $(+)$ sign if both α and β are less than zero, and the $(-)$ sign if one is greater than zero, and the other less than zero.

We take as the definition of multiplication by zero that $\alpha \cdot 0 = 0 \cdot \alpha = 0$. The basic rules of multiplication are at once verified:

$$\alpha\beta = \beta\alpha; \ (\alpha\beta)\gamma = \alpha(\beta\gamma); \ \alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma ,$$

and the product of a given number of factors can be zero if, and only if, at least one of these factors is zero.

Subtraction is defined as the converse operation to addition, i.e. $\alpha - \beta = x$ is equivalent to $x + \beta = \alpha$. Adding $(-\beta)$ to both sides of this equation, we obtain by the properties of addition given above: $x = \alpha + (-\beta)$, i.e. the difference has to be defined in accordance with this formula, so that the operation of subtraction leads to addition. It remains to verify that the expression obtained for x actually satisifies the condition $x + \beta = \alpha$, but this immediately follows from the properties of addition. We shall see how the usual property

is justified, that the inequality $\alpha > \beta$ is equivalent to $\alpha - \beta > 0$. Before this, we turn to division, defining the reciprocal of a given number. If a is a rational number, differing from zero, the number $1/a$ is called the reciprocal. Let α be a real number, not zero. Let $\alpha > 0$ at first, and let a' be any number of $\text{II}(\alpha)$ (it is rational and positive). We define the reciprocal of α by the following section: we put all negative numbers, zero, and $1/a$, in the first class, and the remaining numbers in the second class. Let a certain positive number c_1 belong to the first class of the new section. This means that $c_1 = (1/a_1)$, where a_1' is in $\text{II}(\alpha)$. We take any positive rational number $c_2 < c_1$. It can be written as $c_2 = (1/a_2')$, where a_2' is rational and $a_2' < a_1'$, i.e. a_2' also belongs to $\text{II}(\alpha)$. In other words, if a certain positive number belongs to the first class of the new section, every smaller positive rational number also belongs to the first class. Also, all negative numbers and zero belong there, by hypothesis. Hence it is clear, that in forming the section that defines the reciprocal of α, we maintain the basic hypothesis that any number of the second class is greater than any number of the first class. This number, the reciprocal of α, is denoted by the symbol $1/\alpha$.

If $\alpha < 0$, we define the reciprocal by the formula:

$$\frac{1}{\alpha} = -\frac{1}{|\alpha|}.$$

Using the definition of multiplication, we obtain:

$$\alpha \cdot \frac{1}{\alpha} = 1.$$

We now turn to division. This is the inverse operation to multiplication, i.e. $\alpha : \beta = x$ is equivalent to $x\beta = \alpha$, and, as in subtraction, it is easy to see that, if $\beta \neq 0$, the quotient obtained: $x = \alpha \cdot 1/\beta$ is unique; in this way, division leads to multiplication. Division by zero is impossible.

A number is raised to a positive integral power by multiplication. Extracting a root is defined as the inverse operation to raising to a power. Let α be a real positive number, and n a given integer, greater than unity. We form the following section of the rational numbers: we put in the first class all negative numbers, zero, and all positive numbers whose nth power is less than α, and all remaining numbers go into the second class. Using the definition of multiplication, it is easily seen that the positive number β, defined by this section, satisfies the condition: $\beta^n = \alpha$, i.e. β is the arithmetic value of the root $\sqrt[n]{\alpha}$. If n is even, there will be a second value $(-\beta)$. The root of odd degree of a real negative number is analogously defined (there is a unique answer). Exponential functions will be discussed in full detail later. We now note the following important result: *having once justified the basic laws of operations, all the rules and identities of algebra are simultaneously justified, if letters are understood to represent real numbers.*

42. The strict bounds of numerical sets. Tests for the existence of a limit. We now prove the theorem regarding the strict bounds of a set of real numbers, formulated in [39].

THEOREM. *If a set E of real numbers is bounded above, it has a strict upper bound, and if E is bounded below, it has a strict lower bound.*

We restrict the proof to the first part of the theorem. By hypothesis, all numbers of E are less than a certain number M. We form a section of the real numbers as follows: we put all numbers greater than all numbers of E in the second class, and the remaining real numbers in the first class. We have in the second class, for example, all numbers $(M + p)$, where $p > 0$, and in the first class, for example, all numbers of E. Let β be the real number defined by the section made. By the basic theorem of [40], it will be the greatest in the first class, or the least in the second class. We show that β is in fact the strict upper bound of E. Firstly, there is no number among E greater than β, since all numbers of E are in the first class. Further, there certainly exists a number of E, greater than $(\beta - \varepsilon)$ for any $\varepsilon > 0$, since, if there were no such number, $(\beta - \varepsilon/2)$ would be greater than all numbers of E and would have to belong to the second class, whereas it is actually less than β and is in the first class. The theorem is thus proved. Clearly, if β belongs to E, it will be the greatest of the numbers of E.

We now prove the existence of a limit for a monotonic bounded variable [30]. Let the variable x be continually increasing, or at least, not decreasing, i.e. every one of its values is not less than any previous value. Furthermore, let x be bounded, i.e. there exists a number M, such that all values of x are less than M. We consider the set of all values of x. By the theorem just proved, there exists a strict upper bound β of this set. We show that β is the limit of x. Let ε be an arbitrary positive number. By the definition of strict upper bound, there is a value of x greater than $(\beta - \varepsilon)$. Then, since x is monotonic, all subsequent values of x are greater than $(\beta - \varepsilon)$, whilst on the other hand, they cannot be greater than β; and since ε is arbitrary, it follows that $\beta = \lim x$. The case of a decreasing variable can be worked out in precisely the same way.

We prove a preliminary theorem, before passing to the proof of Cauchy's test [31].

THEOREM. *Given a sequence of finite intervals :*

$$(a_1, b_1), (a_2, b_2), \dots, (a_n, b_n), \dots$$

where each successive interval is contained in the previous one, i.e. $a_{n+1} \geqslant a_n$ *and* $b_{n+1} \leqslant b_n$, *and given also that the lengths of these intervals tend to zero,* i.e. $(b_n - a_n) \to 0$, *then the ends of the intervals, a_n and b_n, tend to a common limit, with increasing n.*

We have by hypothesis: $a_1 \leqslant a_2 \leqslant \dots$, with also $a_n < b_1$ for any value of n. The sequence $a_1, a_2, \dots$ is thus monotonic and bounded, and hence has a limit: $a_n \to a$. Since $(b_n - a_n) \to 0$ by hypothesis, we have: $b_n = a_n + \varepsilon_n$, where $\varepsilon_n \to 0$, and hence b_n also has a limit, equal to a.

We now turn to the proof of Cauchy's test. We confine ourselves to the case of an enumerated variable:

$$x_1, x_2, \dots, x_n, \dots \tag{27}$$

We have to show, that the necessary and sufficient condition for the existence

of a limit of sequence (27) is as follows: for any given positive ε there exists a subscript N, such that

$$| x_m - x_n | < \varepsilon \text{ for } m \text{ and } n > N. \tag{28}$$

We show that the condition is sufficient, i.e. if it is satisfied, sequence (27) has a limit. It follows from our previous discussions [31], that if the condition is satisfied, we can construct a sequence of intervals:

$$(a_1, b_1), (a_2, b_2), \ldots, (a_k, b_k), \ldots$$

with the following properties: each successive interval is contained in the previous one, the length $(b_k - a_k)$ tends to zero, and for every (a_k, b_k) there is a corresponding positive integer N_k, such that all x_s for $s > N_k$ lie in (a_k, b_k). These (a_k, b_k) are the sections $A'_k\, A_k$ of [31]. By the theorem above, we have a common limit:

$$\lim_{k\to\infty} a_k = \lim_{k\to\infty} b_k = a. \tag{29}$$

We show that a is the limit of sequence (27). Let ε be a given positive number. By (29), there exists a positive integer l, such that (a_l, b_l) and all subsequent intervals lie inside the interval $(a - \varepsilon, a + \varepsilon)$.

Hence it follows that all the numbers x_s for $s < N_l$ also belong to this interval, i.e. $| a - x_s | < \varepsilon$ for $s > N_l$. Since ε is arbitrary, we see that a is the limit of sequence (27), and the sufficiency of condition (28) is proved. The necessity of the condition was proved earlier [31]. The proof remains valid for a non-enumerated variable.

43. Properties of continuous functions. Turning to the properties of continuous functions that were given earlier [35], we first prove an auxiliary theorem.

THEOREM 1. *Given $f(x)$ continuous in an interval (a, b), and ε any positive number, the interval can be subdivided into a finite number of new intervals in such a way that $| f(x_2) - f(x_1) | < \varepsilon$, provided x_1 and x_2 belong to the same new interval.*

We shall prove this theorem by *reductio ad absurdum*. We suppose that it is impossible to subdivide (a, b) in the way described. We divide our interval at the centre, obtaining the two intervals:

$$\left(a, \frac{a+b}{2}\right) \text{ and } \left(\frac{a+b}{2}, b\right).$$

If the theorem were true for each of these sub-intervals, it would clearly be true for the whole interval. We must therefore suppose that subvidision in accordance with the theorem is impossible in at least one of the sub-intervals; we take this sub-interval and again divide it into two halves. As before, the theorem is untrue for at least one of these two new halves, and we now divide this half into two, and so on. We thus obtain a sequence of intervals:

$$(a, b), (a_1, b_1), (a_2, b_2), \ldots, (a_n, b_n), \ldots$$

where each successive interval is half of the preceding one, so that the length $(b_n - a_n)$, equal to $(b - a)/2^n$, tends to zero with increasing n. Further, the

theorem is untrue for every (a_n, b_n), i.e. it is not possible to divide any (a_n, b_n) into new intervals so that $|f(x_2) - f(x_1)| < \varepsilon$ provided x_1 and x_2 are in the same new interval. We show that this is absurd.

By the theorem of [42], a_n and b_n have a common limit:

$$\lim a_n = \lim b_n = \alpha, \tag{30}$$

this limit, like all a_n and b_n, lying in (a, b). We first take α in the interior of (a, b). By hypothesis, $f(x)$ is continuous for $x = \alpha$, and hence [34], there is, for the ε given in the theorem, an η such that, for all x in the interval $(\alpha - \eta, \alpha + \eta)$ the inequality is satisfied:

$$|f(\alpha) - f(x)| < \frac{1}{2}\varepsilon. \tag{31}$$

If x_1 and x_2 are any two values from the interval $(\alpha - \eta, \alpha + \eta)$, we have:

$$f(x_2) - f(x_1) = f(x_2) - f(\alpha) + f(\alpha) - f(x_1),$$

whence

$$|f(x_2) - f(x_1)| \leqslant |f(x_2) - f(\alpha)| + |f(\alpha) - f(x_1)|,$$

and by (31):

$$|f(x_2) - f(x_1)| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2},$$

i.e.

$$|f(x_2) - f(x_1)| < \varepsilon \tag{32}$$

for any x_1 and x_2 in $(\alpha - \eta, \alpha + \eta)$. But by (30), there will exist an interval (a_l, b_l) belonging to $(\alpha - \eta, \alpha + \eta)$. Hence inequality (32) will certainly be satisfied for any x_1 and x_2 from this interval, i.e. the theorem is satisfied for (a_l, b_l) even without further subdivision. This contradicts the fact seen above, that the theorem is not satisfied for any (a_n, b_n). The theorem is thus proved, if α is in the interior of (a, b). If, for example, α coincides with the left-hand end of the interval, i.e. $\alpha = a$, the proof is the same, except that the interval $(\alpha, \alpha + \eta)$ is taken instead of $(\alpha - \eta, \alpha + \eta)$.

We now turn to the proof of the third property of [35].

THEOREM 2. *If $f(x)$ is continuous in the interval (a, b), it is uniformly continuous in this interval, i.e. for any given positive ε there exists a positive η, such that $|f(x'') - f(x')| < \varepsilon$ for any x' and x'' of (a, b) that satisfy the inequality $|x'' - x'| < \eta$.*

By Theorem 1, we can subdivide (a, b) into a finite number of new intervals in such a way that $|f(x_2) - f(x_1)| < \varepsilon/2$, provided that x_1 and x_2 belong to the same new interval. Let η be the length of the shortest of the new intervals. We show that our theorem is in fact satisfied for this η. In fact, if x' and x'' are two values from (a, b), satisfying $|x'' - x'| < \eta$, either x' and x'' belong to the same new interval or to two neighbouring intervals. In the first case, we have by the construction of the new intervals: $|f(x'') - f(x')| < \varepsilon/2$, and hence certainly $|f(x'') - f(x')| < \varepsilon$. In the second case, we denote by γ the point of contact of the two adjacent intervals containing x' and x''. We can now write:

$$f(x'') - f(x') = f(x'') - f(\gamma) + f(\gamma) - f(x'),$$

i.e.

$$|f(x'') - f(x')| \leqslant |f(x'') - f(y)| + |f(y) - f(x')| \tag{33}$$

But

$$|f(x'') - f(y)| < \frac{\varepsilon}{2} \quad \text{and} \quad |f(y) - f(x')| < \frac{\varepsilon}{2}, \tag{34}$$

since x'' and y are in the same new interval, and similarly for x' and y. The inequalities (33) and (34) give us $|f(x'') - f(x')| < \varepsilon$, and the theorem is proved.

Theorem 1 also gives us the following corollary.

COROLLARY. *If a function is continuous in an interval (a, b), it is bounded above and below, i.e. simply bounded, in this interval.* In other words, there exists an M, such that $|f(x)| < M$ for all x in (a, b). We take a certain definite ε_0 and let n_0 be the number of new intervals into which (a, b) must be subdivided so as to satisfy Theorem 1 for every $\varepsilon = \varepsilon_0$. We have $|f(x_2) - f(x_1)| < \varepsilon_0$ for any two points belonging to the same new interval. Hence it immediately follows that $|f(x) - f(a)| < n_0\varepsilon_0$ for any x of (a, b), i.e. all $f(x)$ are included between $f(a) - n_0\varepsilon_0$ and $f(a) + n_0\varepsilon_0$.

Since the aggregate of all $f(x)$ of (a, b) is bounded above and below, it has strict upper and lower bounds [42]. We denote the former by β, and the latter by α. We now prove the first property of [35].

THEOREM 3. *A function continuous in an interval (a, b) attains maximum and minimum values in this interval.*

We have to show that there exists an x of (a, b), such that $f(x) = \beta$, and similarly a y, such that $f(x) = \alpha$. We confine ourselves to proving the first assertion, and do so by a contradiction. We suppose that $f(x)$ is not equal to β for any x of (a, b) (and hence is less than β). We form a new function:

$$\varphi(x) = \frac{1}{\beta - f(x)}.$$

Since the denominator is never zero, the new function is also continuous in (a, b) [34]. On the other hand, it follows from the definition of strict upper bound, that for any $\varepsilon > 0$ and $a \leqslant x \leqslant b$ there exists an $f(x)$ lying between $(\beta - \varepsilon)$ and β. Now: $0 < \beta - f(x) < \varepsilon$ and $\varphi(x) < 1/\varepsilon$. Since ε can be taken arbitrarily small, we see that $\varphi(x)$, continuous in (a, b), is not bounded above, which contradicts the above corollary of theorem 1.

We finally prove the second property of [35]. Let $f(x)$ be continuous in (a, b), and k be any number, lying betwen $f(a)$ and $f(b)$. We take for clarity, $f(a) < k < f(b)$. We form a new function:

$$F(x) = f(x) - k,$$

continuous in (a, b). Its values at the ends of the interval are:

$$F(a) = f(a) - k < 0$$

$$F(b) = f(b) - k > 0$$

i.e. $F(x)$ has different signs at the ends of the interval. If we can show that there is an x_0 in (a, b) such that $F(x_0) = 0$, we shall have $f(x_0) - k = 0$,

i.e. $f(x_0) = k$, and the second property is proved. It is thus sufficient to prove the following theorem:

THEOREM 4. *If $f(x)$ is continuous in (a, b), and if $f(a)$ and $f(b)$ are of opposite signs, there exists at least one x_0 for which $f(x_0) = 0$.*

We prove the theorem by supposing it false, as in III. Let $f(x)$ never be zero in (a, b). The new function:

$$\varphi(x) = \frac{1}{f(x)} \tag{35}$$

will also be continuous in (a, b) [34]. We use Theorem 1, and take any $\varepsilon > 0$. We find a finite set of points in (a, b), including its ends, so that the absolute value of the difference of $f(x)$ for any two neighbouring points is less than ε. Noting that $f(a)$ and $f(b)$ have different signs, we can find two neighbouring points ξ_1 and ξ_2 such that the corresponding $f(x)$ have different signs. Thus, whilst $f(\xi_1)$ and $f(\xi_2)$ have different signs, $|f(\xi_2) - f(\xi_1)| < \varepsilon$. Since $f(\xi_1)$ and $f(\xi_2)$ are real numbers, it follows that the absolute value of each is less than ε, i.e. $|f(\xi_1)| < \varepsilon$. But now, by (35), $|\varphi(\xi_1)| > 1/\varepsilon$, and since ε can be taken arbitrarily small, it follows that $\varphi(x)$, continuous in (a, b), is unbounded in the interval. This is absurd. Thus the theorem is proved.

44. Continuity of elementary functions. We proved earlier the continuity of polynomials and rational functions [34]. We now consider the exponential function:

$$y = a^x \quad (a > 0)\,, \tag{36}$$

where we take $a > 1$ for clarity. This function is fully defined for all positive rational x. It is defined for negative x by the formula:

$$a^x = \frac{1}{a^{-x}}\,, \tag{37}$$

with also, $a^0 = 1$. It is thus defined for all rational x. Also, the rules of addition and subtraction of exponents on multiplication and division are known from algebra.

If x is a positive rational number p/q,

$$a^x = \sqrt[q]{a^p},$$

taking the numerical value of the radical. Evidently, $a^p > 1$, and it follows from the definition of a root that $a^x > 1$ for $x > 0$ (taking the definition from [41]). It follows from (37) that $0 < a^x < 1$ for $x < 0$. We now show that $a^{x_2} > a^{x_1}$ if $x_2 > x_1$, i.e. that a^x is an increasing function. In fact,

$$a^{x_2} - a^{x_1} = a^{x_1}(a^{x_2 - x_1} - 1),$$

where $x_2 - x_1 > 0$, so that both factors on the right are positive. We show further that $a^x \to 1$ if $x \to 0$, taking rational values. We first suppose that $x \to 0$, decreasing through all rational values (on the right). Then a^x decreases, but remains greater than unity, and thus has a limit, which we denote by l. Moreover, given this variation of x, $2x$ also tends to zero on the right, through every rational value. We obviously have:

$$a^{2x} = (a^x)^2,$$

so that, passing to the limit:

$$l = l^2 \quad \text{or} \quad l(l-1) = 0\,,$$

i.e. $l = 1$ or $l = 0$. The second possibility is ruled out since $a^x > 1$. Thus, $a^x \to 1$ as $x \to 0$ on the right. It follows from (37) that we have the same limit for $x \to 0$ on the left. Thus in general, $a^x \to 1$ if $x \to 0$, taking rational values. Hence it immediately follows, that if x tends through rational values to a rational limit b, $a^x \to a^b$. In fact,

$$a^x - a^b = a^b\,(a^{x-b} - 1).$$

Here, $(x - b) \to 0$, and by the above, $(a^{x-b} - 1) \to 0$.

We now define function (36) for irrational x. Let α be a certain irrational number, and let $\mathrm{I}(\alpha)$ and $\mathrm{II}(\alpha)$ be the first and second classes respectively of the section in the domain of rational numbers that defines α. We suppose that $x \to \alpha$, increasing and passing through all the rationals of $\mathrm{I}(\alpha)$. Then a^x is increasing but bounded, being in fact less than $a^{x''}$, where x'' is any number of $\mathrm{II}(\alpha)$. Thus, for the above variation of x, a^x has a limit, which we denote for the present by L. Similarly, if $x \to \alpha$, decreasing through the rationals of $\mathrm{II}(\alpha)$, a^x again has a limit. We show that this limit is also L. Let x' belong to $\mathrm{I}(\alpha)$, and x'' to $\mathrm{II}(\alpha)$. We have:

$$a^{x''} - a^{x'} = a^{x'}\,(a^{x''-x'} - 1) < L(a^{x''-x'} - 1),$$

i.e.

$$0 < a^{x''} - a^{x'} < L\,(a^{x''-x'} - 1).$$

For x' and x'' near α, $(x'' - x')$ is as near zero as desired, and the same can be said of $(a^{x''} - a^{x'})$ by the above inequality; whence follows the coincidence of the limits. We take this limit L as defining a^α, i.e. *a^α is the limit to which a^x tends, when $x \to \alpha$ through rational values.* Function (36) is now defined for all real x. It is easy to show from the above, that the function is increasing, i.e. $a^{x_2} > a^{x_1}$, if x_1 and x_2 are any real numbers, satisfying $x_2 > x_1$. We have to consider the cases separately, of x_1 and x_2 both irrational, or of one of them rational. We still have to show that the function is continuous for any real x. We must first show, that $a^x \to 1$ for $x \to 0$, all real values being permitted. The proof is exactly the same as above, for rational x. Also as above, using the formula:

$$a^x - a^\alpha = a^\alpha\,(a^{x-\alpha} - 1),$$

we can show that $a^x \to a^\alpha$ for $x \to \alpha$, which proves that a^x is continuous for any real x.

It is easy to verify that all the basic properties of exponential functions apply for any real exponents. For instance, let α and β be irrationals, and let $x \to \alpha$, $y \to \beta$, x and y varying simultaneously through rational values. We have for rational exponents:

$$a^x\,a^y = a^{x+y}.$$

Passing to the limit, by the continuity of exponential functions proved above, we obtain the same property for irrational exponents:

$$a^\alpha\,a^\beta = a^{\alpha+\beta}.$$

We also obtain the multiplication rule for raising a power to a power:

$$(a^\alpha)^\beta = a^{\alpha\beta},$$

as follows: if $\beta = n$ is a positive integer, the formula follows at once from the rule for adding exponents on multiplication. If $\beta = p/q$ is a positive rational,

$$(a^\alpha)^{p/q} = \sqrt[q]{(a^\alpha)^p} = \sqrt[q]{a^{\alpha p}} = a^{\alpha p/q}.$$

The rule follows at once for negative rationals from (37). We now take β irrational, and let rational r tend to β. By the above:

$$(a^\alpha)^r = a^{\alpha r}.$$

Passing to the limit, and using the continuity of an exponential function, with a^α taken as base on the left, we obtain $(a^\alpha)^\beta = a^{\alpha\beta}$.

Before turning to logarithmic functions, we make a few remarks about inverse functions, which have already been briefly mentioned in the introduction [20]. If $y = f(x)$ is increasing and continuous in (a, b), with $f(a) = A$ and $f(b) = B$, by the second property of continuous functions, when x increases from a to b through all real values, $f(x)$ increases from A to B, passing through all intermediate values. Thus, for every y in (A, B) there is a corresponding definite x in (a, b), and the inverse function $x = \varphi(y)$ will be single-valued and increasing. If $x = x_0$ is inside (a, b), and x runs through a small interval $(x_0 - \varepsilon, x_0 + \varepsilon)$, y will run through a certain interval $(y_0 - \eta_1, y_0 + \eta_2)$, where $y_0 = f(x_0)$. Denoting the least of the two positive η_1 and η_2 by δ, we can say that, if y belongs to $(y_0 - \delta, y_0 + \delta)$, consisting of only part of $(y_0 - \eta_1, y_0 + \eta_2)$, x will belong all the more to the above $(x_0 - \delta, x_0 + \delta)$, i.e. $| \varphi(y) - \varphi(y_0) | < \varepsilon$, provided $| y - y_0 | < \delta$. Since ε is arbitrary, this gives us the continuity of function $x = \varphi(y)$ at $y = y_0$. If x_0 coincides with the end a, for example, we have to take the interval $(x_0, x_0 + \varepsilon)$ in place of $(x_0 - \varepsilon, x_0 + \varepsilon)$ in the above discussion. The working is analogous for $f(x)$ continuous and decreasing.

We return to function (36). Since $a > 1$, $a = 1 + b$, where $b > 0$, and Newton's binomial formula gives for a positive integer $n > 1$:

$$a^n = (1 + b)^n < 1 + nb\ ,$$

whence it is clear that a^x increases indefinitely on indefinite increase of x. Further, it follows from (37) that $a^x \to \infty$ for $x \to \infty$. On noting the above remarks about inverse functions, we can say that the function

$$x = \log_a y\ , \tag{38}$$

the inverse of (36), is single-valued, increasing, and continuous for $y > 0$. The same results are obtained for $0 < a < 1$, except that functions (36) and (38) are decreasing.

We now introduce a new concept, that of a *function of a function*. Let $y = f(x)$ be continuous for $a \leqslant x \leqslant b$, with its values lying in (c, d). Further,

let $z = F(y)$ be continuous in the interval $c \leqslant y \leqslant d$. On taking the above function of x as y, we obtain a function of a function of x:

$$z = F(y) = F(f(x)).$$

We say that this function depends on x through the medium of y. It is defined for $a \leqslant x \leqslant b$. It is easily shown to be continuous in this interval. In fact, an infinitesimal increment of x gives a corresponding infinitesimal increment of y, since $f(x)$ is continuous, and the infinitesimal increment of y gives a corresponding infinitesimal increment of z, since $F(y)$ is continuous.

We now consider the power function

$$z = x^b, \tag{39}$$

with any real exponent b, and taking positive values of x. It immediately follows from the discussion of exponential functions that the value of (39) is defined for every $x > 0$. Using the definition of logarithm and taking natural logarithms, for example, we can write instead of (39):

$$z = e^{b \log x}.$$

Putting $y = b \log x$, and $z = e^y$, we can consider the above as a function of a function of x, and the continuity of (39) for every $x > 0$ follows from the continuity of exponential and logarithmic functions.

The continuity of the trigonometric functions is easily shown by using formulae of elementary trigonometry. From the formula:

$$\sin(x + h) - \sin x = 2 \sin \frac{h}{2} \cos\left(x + \frac{h}{2}\right)$$

we have:

$$|\sin(x + h) - \sin x| \leqslant 2 \left|\sin \frac{h}{2}\right|,$$

since $|\cos(x + h/2)| \leqslant 1$. But for any angle α: $|\sin \alpha| \leqslant |\alpha|$, so that

$$|\sin(x + h) - \sin x| \leqslant |h|$$

and thus the left-hand side tends to zero as $h \to 0$, giving us the continuity of $\sin x$ for all x. The continuity of $\cos x$ for all x is similarly proved. The continuity of $\tan x$ and $\cot x$ for all x, except those for which the denominators in the formulae below become zero, follows at once [34] from:

$$\tan x = \frac{\sin x}{\cos x}, \quad \cot x = \frac{\cos x}{\sin x}.$$

The function $y = \sin x$ is continuous and increasing in the interval $(-\pi/2, \pi/2)$. Using the above remarks concerning inverse functions, we can assert that the principal value of the function $x = \operatorname{arc} \sin y$ [24] is a continuous, increasing function in the interval $-1 \leqslant y \leqslant 1$. The proofs are similar for the continuity of the remaining inverse circular functions.

EXERCISES ON CHAPTER I

1. Prove that, if a and b are real numbers:

$$||a| - |b|| \leqslant |a - b| \leqslant |a| + |b|.$$

2. Prove the following equalities:

$$(a)\quad |ab| = |a| \cdot |b|; \qquad (b)\quad \left|\frac{a}{b}\right| = \frac{|a|}{|b|} \quad (b \neq 0);$$

$$(c)\quad |a|^2 = a^2; \quad (d)\quad \sqrt{a^2} = |a|.$$

3. Determine the range of values of x for which:

$$(a)\quad |x - 1| < 3; \qquad (b)\quad |2x + 1| < 1;$$

$$(c)\quad |x + 1| > 2; \qquad (d)\quad |x - 1| < |x + 1|.$$

4. Find $f(-1), f(0), f(1), f(2), f(3), f(4)$ if $f(x) = x^3 - 6x^2 + 11x - 6$.
5. Find $f(0)$, $f(-\frac{3}{4})$, $f(-x)$, $f(x^{-1})$, $[f(x)]^{-1}$ if $f(x) = \sqrt{1 + x^2}$
6. Let $f(x) = \arccos \log_{10} x$. Find $f(\frac{1}{10})$, $f(1)$, $f(10)$.
7. The function $f(x)$ is linear. Find the function if $f(-1) = 2$ and $f(2) = -3$.
8. Find the formula for the quadratic function $f(x)$ for which $f(0) = 1$, $f(1) = 0$ and $f(3) = 5$.
9. It is given that $f(4) = -2$, $f(5) = 6$. Find the approximate value of $f(4.3)$ by assuming that the function $f(x)$ can be represented by a linear function for $4 \leqslant x \leqslant 5$. *(Linear interpolation formula)*.
10. Find a single formula by means of which to express the function:

$$f(x) = \begin{cases} 0 \text{ if } x \leqslant 0 \\ x \text{ if } x > 0. \end{cases}$$

Define the domains in which the following functions exist:

11. (a) $y = \sqrt{x + 1}$; (b) $y = \sqrt[3]{x + 1}$.
12. $y = 1/(4 - x^2)$.
13. (a) $y = \sqrt{x^2 - 2}$; (b) $y = x\sqrt{x^2 - 2}$.
14. $y = \sqrt{2 + x - x^2}$.
15. $y = \sqrt{-x} + 1/\sqrt{2 + x}$.
16. $y = \sqrt{(x - x^3)}$.
17. $y = \log_{10} [(2 + x)/(2 - x)]$.
18. $y = \log_{10} [(x^2 - 3x + 2)/(x + 1)]$.

19. $y = \arccos\,[2x/(x+1)]$.

20. $y = \arcsin \log_{10}(x/10)$.

21. Define the domain in which the function $y = \sqrt{(\sin 2x)}$ exists.

22. $f(x) = 2x^4 - 3x^3 - 5x^2 + 6x - 10$. Find $\Phi(x) = \frac{1}{2}[f(x) + f(-x)]$ and $\psi(x) = \frac{1}{2}[f(x) - f(-x)]$.

23. Find $\Phi\,[\psi(x)]$ and $\psi\,[\Phi(x)]$, if $\Phi(x) = x^2$ and $\psi(x) = 2^x$.

24. Find $f\{f\,[f(x)]\}$ if $f(x) = 1/(1-x)$.

25. Find $f(x+1)$ if $f(x-1) = x^2$.

26. Let $f(n)$ be the sum of n terms of an arithmetical progression. Show that $f(n+3) - 3f(n+2) + 3f(n+1) - f(n) = 0$.

27. Prove that if $f(x)$ is an exponential function, i.e. $f(x) = a^x$, $x > 0$, and if x_1, x_2, x_3 form an arithmetical progression, then $f(x_1), f(x_2), f(x_3)$ form a geometrical progression.

28. Let $f(x) = \log\,[(1+x)/(1-x)]$. Show that

$$f(x) + f(y) = f\left(\frac{x+y}{1+xy}\right).$$

29. Let $\varphi(x) = \frac{1}{2}(a^x + a^{-x})$ and $\psi(x) = \frac{1}{2}(a^x - a^{-x})$. Show that $\varphi(x+y) = \varphi(x)\,\varphi(y) + \psi(x)\,\psi(y)$ and $\psi(x+y) = \varphi(x)\,\psi(y) + \varphi(y)\,\psi(x)$.

30. Find $f(-1), f(0), f(1)$ if

$$f(x) = \begin{cases} \arcsin\ x \text{ for } -\infty \leqslant x \leqslant 0 \\ \arctan\ x \text{ for } 0 < x < +\infty \end{cases},$$

31. Find the zeros and the domains in which the following functions are respectively positive and negative:

(*a*) $y = 1 + x$; (*b*) $y = 2 + x - x^2$; (*c*) $y = 1 - x + x^2$;

(*d*) $y = x^3 - 3x$; (*e*) $y = \log_{10}[2x/(1+x)]$.

32. Find the inverses of the following functions:

(*a*) $y = 2x + 3$; (*b*) $y = x^2 - 1$; (*c*) $y = \sqrt[3]{1 - x^3}$;

(*d*) $y = \log_{10}(\frac{1}{2}x)$; (*e*) $y = \arctan 3x$.

33. Find the inverse of the function defined by:

$$y = \begin{cases} x \text{ if } x \leqslant 0 \\ x^2 \text{ if } x > 0 \end{cases}$$

34. Write the following functional relationships in the form of a chain of equations the elements of which are simple functions (e.g. $y = \left(1+\frac{1}{x}\right)^n$ could be written as $y = u^n$, $u = 1 + v$, $v = \frac{1}{x}$):
(*a*) $y = (2x - 5)^{10}$; (*b*) $y = 2^{\cos x}$; (*c*) $y = \log_{10}\tan\frac{1}{2}x$;
(*d*) $y = \arcsin(3^{-x})$.

35. From each of the following chains of equations form a compound function written in the form of a single equation:
(*a*) $y = u^2$, $u = \sin x$; (*b*) $y = \arctan u$, $u = \sqrt{v}$, $v = \log_{10}x$;
(*c*) $y = 2u$ if $u \leqslant 0$, $y = 0$ if $u > 0$, $u = x^2 - 1$.

36. Write down in explicit form the functions y satisfying the following equations.
(*a*) $x^2 - \arccos y = \pi$; (*b*) $10^x + 10^y = 10$; (*c*) $x + |y| = 2y$.
Find the domain of definition of each of the functions.

Construct the graphs of the bilinear functions (hyperbolas):

37. $y = 1/x$.
38. $y = 1/(1 - x)$.
39. $y = (x - 2)/(x + 2)$.
40. $y = y_0 + m/(x - x_0)$ if $x_0 = 1$, $y_0 = -1$, $m = 6$.
41. $y = (2x - 3)/(3x + 2)$.

Construct the graphs of the rational functions:

42. $y = x + 1/x$. **43.** $y = x^2/(x + 1)$. **44.** $y = 1/x^2$. **45.** $y = 1/x^3$.
46. $y = 10/(x^2 + 1)$ *(Witch of Agnesi)*. **47.** $y = 2x/(x^2 + 1)$.
48. $y = (x + 1)/x^2$. **49.** $y = (x^2 + 1)/x$.

Construct the graphs of the trigonometric functions:

50. $y = \sin x$. **51.** $y = \cos x$. **52.** $y = \tan x$. **53.** $y = \cot x$.
54. $y = \sec x$. **55.** $y = \operatorname{cosec} x$. **56.** $y = A \sin x$ with $A = 1, 10, 1/2, -2$.
57. $y = \sin nx$ with $n = 1, 2, 3, \frac{1}{2}$. **58.** $y = \sin(x - \varphi)$, with $\varphi = 0, \frac{1}{2}\pi, \frac{3}{2}\pi, \pi, -\frac{1}{4}\pi$. **59.** $y = a \sin x + b \cos x$, with $a = 6$, $b = -8$. **60.** $y = \sin x + \cos x$. **61.** $y = \cos^2 x$. **62.** $y = x + \sin x$.
63. $y = x \sin x$.

Construct the graphs of the exponential and logarithmic functions:

64. $y = a^x$ with $a = 2, \frac{1}{2}, e\ (= 2.718\ldots)$.
65. $y = \log_a x$, with $a = 10, 2, \frac{1}{2}, e$.
66. $y = \sinh x = \frac{1}{2}(e^x - e^{-x})$. **67.** $y = \cosh x = \frac{1}{2}(e^x + e^{-x})$.
68. $y = \tanh x \equiv \sinh x/\cosh x$. **69.** $y = 10^{1/x}$.

70. $y = e^{-x^2}$ *(error curve)*. **71.** $y = 2^{-1/x^2}$. **72.** $y = \log_{10}x^2$.
73. $y = (\log_{10}x)^2$. **74.** $y = \log_{10}(\log_{10}x)$. **75.** $y = 1/\log_{10}x$.
76. $y = \log_{10}(1/x)$. **77.** $y = \log_{10}(-x)$. **78.** $y = \log_2(1 + x)$.
79. $y = \log_{10}(\cos x)$. **80.** $y = 2^{-x} \sin x$.

81. Prove that as $n \to \infty$ the limit of the sequence

$$1, \frac{1}{4}, \frac{1}{9}, \ldots, \frac{1}{n^2}, \ldots$$

is equal to zero. For what values of n will we have the inequality $n^{-2} < \varepsilon$ where ε is an arbitrary positive number.
Give an estimate of the lowest such n for (*a*) $\varepsilon = 0.1$; (*b*) $\varepsilon = 0.01$; (*c*) $\varepsilon = 0.001$.

82. Prove that the limit of the sequence $x_n = n/(n + 1)$, $(n = 1, 2, \ldots)$ as $n \to \infty$ is equal to 1. If ε is an arbitrary positive number find N such that $|x_n - 1| < \varepsilon$ for $n > N$. Find N if (*a*) $\varepsilon = 0.1$; (*b*) $\varepsilon = 0.01$; (*c*) $\varepsilon = 0.001$.

83. Prove that

$$\lim_{x \to 2} x^2 = 4.$$

Given a positive number ε find a positive number δ such that $|x^2 - 4| < \varepsilon$ whenever $|x - 2| < \delta$. Calculate δ if (*a*) $\varepsilon = 0.1$; (*b*) $\varepsilon = 0.01$; (*c*) $\varepsilon = 0.001$.

84. Write out precisely what is meant by the following statements:
(*a*) $\lim\limits_{x \to +0} \log_{10}x = -\infty$; (*b*) $\lim\limits_{x \to \infty} 2^x = +\infty$; (*c*) $\lim\limits_{x \to \infty} f(x) = \infty$.

85. Find the limits of the sequences:

(*a*) $1, -\frac{1}{2}, \frac{1}{3}, -\frac{1}{4}, \ldots, \frac{(-1)^{n-1}}{n}, \ldots$;

(*b*) $\frac{2}{1}, \frac{4}{3}, \frac{6}{5}, \ldots, \frac{2n}{2n-1}, \ldots$;

(*c*) $\sqrt{2}, \sqrt{2\sqrt{2}}, \sqrt{2\sqrt{2\sqrt{2}}}, \ldots$;

(*d*) 0.2, 0.23, 0.233, 0.2333, ...

86. $\lim\limits_{n \to \infty} \left(\frac{1}{n^2} + \frac{2}{n^2} + \frac{3}{n^2} + \ldots + \frac{n-1}{n^2}\right)$.

87. $\lim\limits_{n \to \infty} \frac{1}{n^3}(n + 1)(n + 2)(n + 3)$.

88. $\lim\limits_{n \to \infty} \left[\frac{1 + 3 + 5 + 7 + \ldots + (2n - 1)}{n + 1} - \frac{2n + 1}{2}\right]$.

89. $\lim\limits_{n\to\infty} \frac{n+(-1)^n}{n-(-1)^n}$.

90. $\lim\limits_{n\to\infty} \frac{2^{n+1}+3^{n+1}}{2^n+3^n}$.

91. $\lim\limits_{n\to\infty} \left(\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots+\frac{1}{2^n}\right)$.

92. $\lim\limits_{n\to\infty} \left[1-\frac{1}{3}+\frac{1}{9}-\frac{1}{27}+\ldots+\frac{(-1)^{n-1}}{3^{n-1}}\right]$

93. $\lim\limits_{n\to\infty} \frac{1^2+2^2+3^2+\ldots+n^2}{n^3}$

94. $\lim\limits_{n\to\infty} (\sqrt{n+1}-\sqrt{n})$.

95. $\lim\limits_{n\to\infty} \frac{n \sin n!}{n^2+1}$.

96. $\lim\limits_{x\to\infty} \frac{(x+1)^2}{x^2+1}$

97. $\lim\limits_{x\to\infty} \frac{1000x}{x^2-1}$

98. $\lim\limits_{x\to\infty} \frac{x^2-5x+1}{3x+7}$

99. $\lim\limits_{x\to\infty} \frac{2x^2-x+3}{x^3-8x+5}$

100. $\lim\limits_{x\to\infty} \frac{\sqrt{x}}{\sqrt{x+\sqrt{x+\sqrt{x}}}}$

101. $\lim\limits_{x\to 1} \frac{\sqrt{x-1}}{x-1}$

102. $\lim\limits_{x\to 64} \frac{\sqrt{x}+8}{\sqrt[3]{x}-4}$

103. $\lim\limits_{x\to 1} \frac{\sqrt[3]{x}-1}{\sqrt[4]{x}-1}$.

104. $\lim\limits_{x\to 1} \frac{\sqrt[3]{x^2}-2\sqrt[3]{x}+1}{(x-1)^2}$

105. $\lim\limits_{x\to 7} \frac{2-\sqrt{x-3}}{x^2-49}$

106. $\lim\limits_{x\to\infty} \frac{x-8}{\sqrt[3]{x}-2}$

107. $\lim\limits_{x\to 1} \frac{\sqrt{x}-1}{\sqrt[3]{x}-1}$

108. $\lim\limits_{x\to 4} \frac{3-\sqrt{5+x}}{1-\sqrt{5-x}}$

109. $\lim\limits_{x\to 0} \frac{\sqrt{(1+x)}-\sqrt{(1-x)}}{x}$

110. $\lim\limits_{h\to 0} \frac{\sqrt{x+h}-\sqrt{x}}{h}$

111. $\lim\limits_{h\to 0} \frac{\sqrt[3]{x+h}-\sqrt[3]{x}}{h}$

112. $\lim\limits_{x\to\infty} (\sqrt{x+a}-\sqrt{x})$

113. $\lim\limits_{x\to\infty} [\sqrt{x(x+a)}-x]$

114. $\lim\limits_{x\to\infty} [\sqrt{x^2-5x+6}-x]$

115. $\lim\limits_{x\to\infty} x[\sqrt{x^2+1}-x]$

116. $\lim\limits_{x\to\infty}[x+\sqrt[3]{1-x^3}]$

117. (a) $\lim\limits_{x\to 2}\frac{\sin x}{x}$; (b) $\lim\limits_{x\to\infty}\frac{\sin x}{x}$

118. $\lim\limits_{x\to 0}\frac{\sin 3x}{x}$

119. $\lim\limits_{x\to 0}\frac{\sin 5x}{\sin 2x}$

120. $\lim\limits_{x\to 1}\frac{\sin \pi x}{\sin 3\pi x}$

121. $\lim\limits_{n\to\infty} n\sin\frac{\pi}{n}$

122. $\lim\limits_{x\to 0}\frac{1-\cos x}{x^2}$

123. $\lim\limits_{x\to a}\frac{\sin x-\sin a}{x-a}$

124. $\lim\limits_{x\to a}\frac{\cos x-\cos a}{x-a}$

125. $\frac{1-\sin\frac{1}{2}x}{\pi-x}$

126. $\lim\limits_{x\to\frac{1}{3}\pi}\frac{1-2\cos x}{\pi-3x}$

127. $\lim\limits_{x\to 0}\frac{\sqrt{1+\sin x}-\sqrt{1-\sin x}}{x}$

128. $\lim\limits_{n\to\infty}\left(1-\frac{1}{n}\right)^n$

129. $\lim\limits_{x\to\infty}\left(1+\frac{2}{x}\right)^x$

130. $\lim\limits_{x\to\infty}\left(\frac{x}{x+1}\right)^x$

131. $\lim\limits_{x\to\infty}\left(\frac{x-1}{x+3}\right)^{x+2}$

132. $\lim\limits_{n\to\infty}\left(1+\frac{x}{n}\right)^n$

133. $\lim\limits_{x\to\infty}(1+\sin x)^{1/x}$

134. Show that the function $y = x^2$ is continuous for any value of the argument x.

135. Prove that the polynomial $P(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$ is continuous for all values of x.

136. Prove that the rational function $R(x) = (a_0x^n + a_1x^{n-1} + \ldots + a_n)/(b_0x^m + b_1x^{m-1} + \ldots + b_m)$ is continuous for all values of x except those for which the denominator vanishes.

137. Prove that the function $y = \sqrt{x}$ is continuous for $x \geqslant 0$.

138. Prove that if the function $f(x)$ is continuous and non-negative in the interval (a, b) the function $F(x) = \sqrt{f(x)}$ is continuous in that interval.

139. Prove that the function $y = \cos x\xi$ is continuous for all values of x.

140. For what values of x are the functions (a) $\tan x$, (b) $\cot x$ continuous?

141. Prove that the function $y = |x|$ is continuous for all x. Construct the graph of this function.

142. Prove that the absolute value of a continuous function is a continuous function.

143. A function is given by the formula:

$$f(x) = \begin{cases} (x^2 - 4)/(x - 2) & \text{if } x \neq 2 \\ A & \text{if } x = 2 \end{cases}$$

How should we choose the value of $A = f(2)$ in order that the function $f(x)$ will be continuous at $x = 2$? Sketch the graph of the function $y = f(x)$.

144. The function $f(x)$ is not defined for $x = 0$. Define $f(0)$ in order that $f(x)$ will be continuous at $x = 0$ if:
(*a*) $f(x) = \frac{1}{x}[(1 + x)^n - 1]$; (*b*) $f(x) = (1 - \cos x)/x^2$;
(*c*) $f(x) = \frac{1}{x}[\log(1 + x) - \log(1 - x)]$; (*d*) $f(x) = \frac{1}{x}(e^x - e^{-x})$;
(*e*) $f(x) = x^2 \sin(1/x)$; (*f*) $f(x) = x \cot x$.

CHAPTER II

DIFFERENTIATION: THEORY AND APPLICATIONS

§ 3. Derivatives and differentials of the first order

45. The concept of derivative. We consider a point moving in a straight line. The path s traversed by the point, measured from some definite point of the line, is evidently a function of time t:

$$s = f(t).$$

A corresponding value of s is defined for every definite value of t. If t receives an increment Δt, the path $s + \Delta s$ will then correspond to the new instant $t + \Delta t$, where Δs is the path traversed in the interval Δt. In the case of uniform motion, the increment of path is proportional to the increment of time, and the ratio $\Delta s/\Delta t$ represents the constant velocity of the motion. This ratio is in general dependent both on the choice of the instant t and on the increment Δt, and represents the *average velocity* of the motion during the interval from t to $t + \Delta t$. This average velocity is the velocity of an imaginary point which moves uniformly and traverses path Δs in time Δt. For example, we have in the case of uniformly accelerated motion:

$$s = \frac{1}{2} gt^2 + v_0 t$$

and

$$\frac{\Delta s}{\Delta t} = \frac{\frac{1}{2} g(t+\Delta t)^2 + v_0(t+\Delta t) - \frac{1}{2} gt^2 - v_0 t}{\Delta t} = gt + v_0 + \frac{1}{2} g\Delta t.$$

The smaller the interval of time t, the more we are justified in taking the motion of the point in question as uniform in this interval, and the limit of the ratio $\Delta s/\Delta t$, with Δt tending to zero, defines *the velocity v at the given instant t:*

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t}.$$

Thus, in the case of uniformly accelerated motion:

$$v = \lim_{\Delta t \to 0} \frac{\Delta s}{\Delta t} = \lim_{\Delta t \to 0} \left(gt + v_0 + \frac{1}{2} g \, \Delta t \right) = gt + v_0.$$

The velocity v, like the path s, is a function of t; this function is called the *derivative* of function $f(t)$ with respect to t; thus, *the velocity is the derivative of the path with respect to time.*

Suppose that a substance takes part in a chemical reaction. The quantity x of this substance, taking part in the reaction at the instant t, is a function of t. There is a corresponding increment Δx of magnitude x for an increment of time Δt, and the ratio $\Delta x / \Delta t$ gives the *average speed of the reaction* in the interval Δt, whilst the limit of this ratio as Δt tends to zero gives the *speed of the chemical reaction at the given instant* t.

We considered above [32] the quantity of heat Q, absorbed by a body, as a function of its temperature $t°$. Let $\Delta t°$ and ΔQ be the corresponding increments of temperature and quantity of heat. Accurate measurements indicate that ΔQ is not proportional to $\Delta t°$, and the ratio $\Delta Q / \Delta t$ gives the so-called *average specific heat* of the body in the temperature interval from $t°$ to $t° + \Delta t°$, whilst the limit of this ratio as $\Delta t°$ tends to zero gives the specific heat of the body at $t°$, this being the derivative of the quantity of heat with respect to temperature.

The above examples lead us to the following concept of the derivative of a function:

The derivative of a given function $y = f(x)$ is defined as the limit of the ratio of the increment Δy of the function to the corresponding increment Δx of the independent variable, when the latter tends to zero.

The symbols y' or $f'(x)$ are used to denote the derivative:

$$y' = f'(x) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

The operation of finding the derivative is called *differentiation*.

It is possible for the above limit not to exist, in which case the derivative does not exist. Assuming that the derivative exists, we can write:

$$\frac{f(x + \Delta x) - f(x)}{\Delta x} = f'(x) + \alpha,$$

where $\alpha \to 0$, as $\Delta x \to 0$ [27].

We now have:

$$f(x + \Delta x) - f(x) = [f'(x) + a]\Delta x\,,$$

whence it is immediately clear that $[f(x + \Delta x) - f(x)] \to 0$ if $\Delta x \to 0$, i.e. *if the derivative exists for some value of x, the function is continuous for this value of x.* The converse statement does not hold, i.e. nothing can be said about the existence of a derivative from the continuity of a function. We note that in finding the derivative we take the fraction $\Delta y/\Delta x$, with the numerator and denominator both tending to zero; but we suppose that Δx never in fact becomes zero.

46. Geometrical significance of the derivative. We turn to the graph of the function $y = f(x)$ to see the geometrical meaning of the derivative. We take a point M of the graph with coordinates (x, y), and an adjacent point N of the graph with coordinates $(x + \Delta x, y + \Delta y)$. We draw the ordinates $\overline{M_1 M}$ and $\overline{N_1 N}$ of these points, and take a line through M parallel to axis OX. We evidently have (Fig. 50):

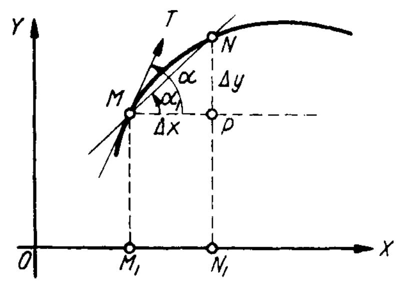

FIG. 50

$$\overline{MP} = \overline{M_1 N_1} = \Delta x, \quad \overline{M_1 M} = y, \quad \overline{N_1 N} = y + \Delta y, \quad \overline{PN} = \Delta y. \qquad (1)$$

The ratio $\Delta y/\Delta x$ is clearly equal to the tangent of the angle a_1 that MN forms with the positive direction of OX. As Δx tends to zero, point N will tend to point M, whilst remaining on the curve; MN becomes, in the limiting position, the *tangent* MT to the curve at the point M; hence, the derivative $f'(x)$ is *equal to the tangent of the angle a formed by the tangent to the curve at the point* $M(x, y)$ *with the positive direction of axis OX*, i.e. *is equal to the slope of this tangent.*

Attention must be paid to the rule of signs when working out the segments in accordance with formula (1), remembering that increments Δx and Δy can be either negative or positive.

We see that the existence of a derivative $f'(x)$ is bound up with the existence of a tangent to the curve corresponding to $y = f(x)$. A continuous curve may have no tangent at all at certain points, or it may have a tangent parallel to axis OY, with infinitely large

slope (Fig. 51); the function $f(x)$ then has no derivative for the corresponding value of x.

A curve can have any number of such singular points, and it is even possible to construct a continuous function, as may be shown, such that it has no derivative for any value of x. The curve corresponding to this function cannot be represented geometrically.

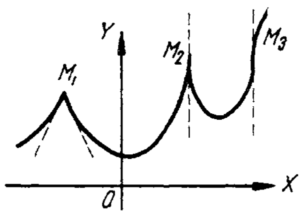

FIG. 51

Denoting for simplicity the increment of the independent variable by h, we have the ratio:

$$\frac{f(x+h)-f(x)}{h}. \tag{2}$$

If the number x is fixed in the interval in which $f(x)$ is defined, the ratio (2) is a function of h, defined for all h sufficiently close to zero, except $h = 0$. The limit of this ratio as $h \to 0$ has to be determined in accordance with what was said in [32]. If the limit exists, it gives us the derivative $f'(x)$. The existence of the limit is equivalent to the following [32]. For any given positive ε there exists a positive η such that

$$\left|f'(x) - \frac{f(x+h)-f(x)}{h}\right| < \varepsilon \text{ for } |h| < \eta \text{ and } h \neq 0.$$

It can happen that ratio (2) has a limit for h tending to zero on the side of positive values (on the right), and on the side of negative values (on the left). These limits are usually denoted by $f'(x+0)$ and $f'(x-0)$, being called respectively *the derivative on the right* and *the derivative on the left*. If these limits differ, they give the slopes of the tangents to the curve at its bend point (if the tangents exist). Figure 51 shows these tangents at the point M_1.

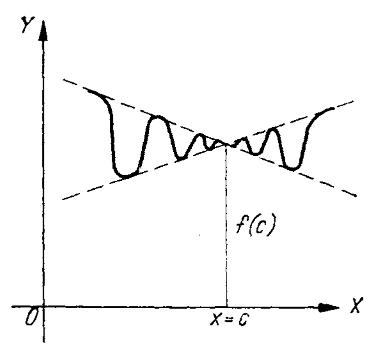

FIG. 52

The existence of the derivative is equivalent to the existence of derivatives $f'(x+0)$ and $f'(x-0)$ and to the fact of these being equal, so that we have $f'(x) = f'(x+0) = f'(x-0)$. It is possible for a continuous function to have points for which there is neither derivative $f'(x+0)$ nor $f'(x-0)$. Such a curve is shown in Fig. 52. It has neither derivative for $x = c$.

If a continuous function is given solely in an interval (a, b), we can only form the derivative on the right, $f'(a + 0)$, at $x = a$, and only the derivative on the left, $f'(b - 0)$, at $x = b$. When $f(x)$ is said to have *a derivative* $f'(x)$ *in the (closed) interval* (a, b), this must be taken to mean the derivative in the ordinary sense for interior points of the interval, and in the special sense indicated at the ends of the interval.

If $f(x)$ is defined in the interval (A, B), wider than (a, b), i.e. $A < a$ and $B > b$, and has the ordinary derivative $f'(x)$ inside (A, B) it will certainly have a derivative in the sense indicated over (a, b).

47. Derivatives of some simple functions. It follows from the concept of derivative that, *to find the derivative, the increment given to the function must be divided by the corresponding increment of the independent variable, the limit of their ratio then being found as the increment of the independent variable tends to zero.* We use this rule for some elementary functions.

I. $y = b$ (constant) [12].

$$y' = \lim_{h\to 0} \frac{b-b}{h} = \lim_{h\to 0} 0 = 0,$$

i.e. *the derivative of a constant is zero.*

II. $y = x^n$ (n a positive integer).

$$y' = \lim_{h\to 0} \frac{(x+h)^n - x^n}{h} =$$

$$= \lim_{h\to 0} \frac{x^n + nhx^{n-1} + \frac{n(n-1)}{2!} h^2 x^{n-2} + \ldots + h^n - x^n}{h} =$$

$$= \lim_{h\to 0} nx^{n-1} + \frac{n(n-1)}{2!} hx^{n-2} + \ldots + h^{n-1} = nx^{n-1}.$$

In particular, if $y = x$, $y' = 1$. We later generalize this rule for differentiation of a power function for any value of the exponent n.

III. $y = \sin x$.

$$y' = \lim_{h\to 0} \frac{\sin(x+h) - \sin x}{h} = \lim_{h\to 0} \frac{2\cos\left(x + \frac{1}{2}h\right)\sin\frac{1}{2}h}{h} =$$

$$= \lim_{h\to 0} \cos\left(x + \frac{1}{2}h\right) \frac{\sin\frac{1}{2}h}{\frac{1}{2}h} = \cos x,$$

since $\dfrac{\sin\frac{1}{2}h}{\frac{1}{2}h} \to 1$ for $\frac{1}{2}h \to 0$ [33].

IV. $y = \cos x$.

$$y' = \lim_{h\to 0} \frac{\cos(x+h) - \cos x}{h} = \lim_{h\to 0} - \frac{2\sin\left(x + \frac{1}{2}h\right)\sin\frac{1}{2}h}{h} =$$

$$= -\lim \sin\left(x + \frac{1}{2}h\right)\frac{\sin\frac{1}{2}h}{\frac{1}{2}h} = -\sin x.$$

V. $y = \log x$ $(x > 0)$.

$$y' = \lim_{h\to 0} \frac{\log(x+h) - \log x}{h} = \lim_{h\to 0} \frac{\log\left(1 + \frac{h}{x}\right)}{h} =$$

$$= \lim_{h\to 0} \frac{1}{x} \frac{\log\left(1 + \frac{h}{x}\right)}{\frac{h}{x}} = \frac{1}{x},$$

since for $h \to 0$, $a = h/x$ also tends to 0, and $\log(1 + a)/a \to 1$ [38].

VI. $y = cu(x)$, where c is a constant, and $u(x)$ a function of x.

$$y' = \lim_{h\to 0} \frac{cu(x+h) - cu(x)}{h} = c\lim_{h\to 0} \frac{u(x+h) - u(x)}{h} = cu'(x),$$

i.e. *the derivative of the product of a constant and a variable is equal to the product of the constant and the derivative of the variable*, or, in other words, *the constant can be taken outside the sign of the derivative.*

VII. $y = \log_a x$.

As we know, $\log_a x = \log x \times 1/\log a$ [38]. Using Rule VI, we obtain:

$$y' = \frac{1}{x} \times \frac{1}{\log a}.$$

VIII. We consider the derivative of the sum of several variables; we confine ourselves to three terms for clarity:

$$y = u(x) + v(x) + w(x),$$

$$y' = \lim_{h\to 0} \frac{[u(x+h) + v(x+h) + w(x+h)] - [u(x) + v(x) + w(x)]}{h} =$$

$$= \lim_{h\to 0} \left[\frac{u(x+h) - u(x)}{h} + \frac{v(x+h) - v(x)}{h} + \frac{w(x+h) - w(x)}{h}\right] =$$

$$= u'(x) + v'(x) + w'(x),$$

i.e. *the derivative of the sum of any given number of functions is equal to the sum of the derivatives of these functions.*

IX. We now consider the derivative of the product of two functions: $y = u(x) \times v(x)$,

$$y' = \lim_{h \to 0} \frac{u(x+h) \times v(x+h) - u(x) \times v(x)}{h}.$$

Adding and subtracting $u(x+h) \times v(x)$ in the numerator, then re-arranging, we get:

$$y' = \lim_{h \to 0} \frac{u(x+h) \times v(x+h) - u(x+h) \times v(x) + u(x+h) \times v(x) - u(x) \times v(x)}{h} =$$

$$= \lim_{h \to 0} u(x+h) \frac{v(x+h) - v(x)}{h} + \lim_{h \to 0} v(x) \frac{u(x+h) - u(x)}{h} =$$

$$= u(x) \times v'(x) + v(x) \times u'(x),$$

i.e. we have shown that for two factors, *the derivative of the product is equal to the sum of the products of each factor with the derivative of the other.*

We prove the applicability of this rule to three factors by combining two factors in one group and using the rule for two:

$$y = u(x) \times v(x) \times w(x),$$

$$y' = \{[u(x) \times v(x)] \times w(x)\}' = [u(x) \times v(x)] \times w'(x) +$$
$$+ w(x) \times [u(x) \times v(x)]' =$$
$$= u(x) \times v(x) \times w'(x) + u(x) \times v'(x) \times w(x) + u'(x) \times v(x) \times w(x).$$

Using the well-known method of mathematical induction, the rule for two can easily be extended to the case of any finite number of factors.

X. Now let y be a quotient:

$$y = \frac{u(x)}{v(x)}, \quad y' = \lim_{h \to 0} \frac{\frac{u(x+h)}{v(x+h)} - \frac{u(x)}{v(x)}}{h} =$$

$$= \lim_{h \to 0} \frac{1}{v(x)\, v(x+h)} \cdot \frac{u(x+h)\, v(x) - v(x+h)\, u(x)}{h}.$$

Adding and subtracting $u(x)v(x)$ in the numerator of the second fraction, and taking the continuity of $v(x)$ into account, we get:

$$y' = \lim_{h \to 0} \frac{1}{v(x)\, v(x+h)} \times \frac{u(x+h)\, v(x) - u(x)\, v(x) + u(x)\, v(x) - v(x+h)\, u(x)}{h} =$$

$$= \lim_{h \to 0} \frac{1}{v(x)\, v(x+h)} \left[v(x) \frac{u(x+h) - u(x)}{h} - u(x) \frac{v(x+h) - v(x)}{h} \right] =$$

$$= \frac{u'(x)\, v(x) - v'(x)\, u(x)}{[v(x)]^2},$$

i.e. *the derivative of a fraction (quotient) is equal to the denominator times the derivative of the numerator, minus the numerator times the derivative of the denominator, all divided by the square of the denominator.*

XI. $y = \tan x$.

$$y' = \left(\frac{\sin x}{\cos x}\right)' = \frac{(\sin x)' \cos x - (\cos x)' \sin x}{\cos^2 x} =$$

$$= \frac{\cos^2 x + \sin^2 x}{\cos^2 x} = \frac{1}{\cos^2 x}.$$

XII. $y = \cot x$.

$$y' = \left(\frac{\cos x}{\sin x}\right)' = \frac{(\cos x)' \sin x - (\sin x)' \cos x}{\sin^2 x} =$$

$$= \frac{-\sin^2 x - \cos^2 x}{\sin^2 x} = -\frac{1}{\sin^2 x}.$$

In deducing Rules VI, VIII, IX and X, we assumed that the functions $u(x)$, $v(x)$, $w(x)$ have derivatives, and proved the existence of a derivative of the function y.

48. Derivatives of functions of a function, and of inverse functions. We recall the concept of function of a function [44]. Let $y = f(x)$ be a function, continuous in some interval $a \leqslant x \leqslant b$, with its value lying in the interval $c \leqslant y \leqslant d$. Further, let $z = F(y)$ be a function continuous in the interval $c \leqslant y \leqslant d$. Taking the above function of x as y, we obtain a function of a function (composite function) of x:

$$z = F(y) = F(f(x)).$$

This function is said to depend on x expressed by means of y. It is easily seen that the function is continuous in the interval $a \leqslant x \leqslant b$. In fact, given an infinitesimal increment of x, there is a corresponding infinitesimal increment of y, due to the continuity of $f(x)$, and given the infinitesimal increment of y, there is a corresponding infinitesimal increment of z, due to the continuity of $F(y)$.

We make one remark, before deducing the rule for differentiation of a function of a function. If $z = F(y)$ has a derivative for $y = y_0$, it follows from the above [45], that we can write:

$$\Delta z = F(y_0 + \Delta y) - F(y_0) = [F'(y_0) + a]\, \Delta y\,, \tag{3}$$

where a is a function of Δy, defined for all positive values of Δy approaching zero, and where $a \to 0$ as $\Delta y \to 0$ $(\Delta y \neq 0)$. Equation (3) remains valid for $\Delta y = 0$ with any choice of a, since for $\Delta y = 0$, Δz

also $= 0$. It is natural from the above to take $a = 0$ for $\Delta y = 0$. Having agreed on this, we can take $a \to 0$ in formula (3) for $\Delta y \to 0$ in any manner, even whilst taking values equal to zero. We now formulate the theorem about the derivative of a function of a function.

THEOREM. *If $y = f(x)$ has a derivative $f'(x_0)$ at $x = x_0$ and $z = F(y)$ has a derivative $F'(y_0)$ at $y_0 = f(x_0)$, the function of a function $F(f(x))$ has a derivative at $x = x_0$ equal to the product $F'(y_0)f'(x_0)$.*

Let Δx be the increment (not zero) that we give to the value x_0 of the independent variable x, and $\Delta y = f(x_0 + \Delta x) - f(x_0)$ be the corresponding increment of variable y (its value can be zero). Further, let $\Delta z = F(y_0 + \Delta y) - F(y_0)$. The derivative of the function of a function $z = F(f(x))$ with respect to x, at $x = x_0$, is evidently equal to the limit of the ratio $\Delta z/\Delta x$ as $\Delta x \to 0$, if this limit exists. We divide both sides of (3) by Δx:

$$\frac{\Delta z}{\Delta x} = [F'(y_0) + a]\frac{\Delta y}{\Delta x}.$$

As $\Delta x \to 0$, Δy also $\to 0$, due to the continuity of function $y = f(x)$ at the point $x = x_0$; and hence, as we have shown above, $a \to 0$. The ratio $\Delta y/\Delta x$ now tends to the derivative $f'(x_0)$, and on passing to the limit in the equation above, we obtain:

$$\lim_{\Delta x \to 0} \frac{\Delta z}{\Delta x} = F'(y_0) f'(x_0),$$

which proves the theorem. We remark that the continuity of $f(x)$ at $x = x_0$ follows from the assumption of the existence of a derivative $f'(x_0)$ [45].

This theorem can be put in the following form, as a rule for differentiation of functions of a function: *the derivative of a function of a function is equal to the product of the derivative with respect to the intermediate variable and the derivative of the intermediate variable with respect to the independent variable :*

$$z'_x = F'(y)f'(x).$$

We pass to the rule for differentiation of inverse functions. If $y = f(x)$ is continuous and increasing in the interval (a, b) (i.e. the greatest value of y corresponds to the greatest value of x), with $A = f(a)$ and $B = f(b)$, we know, [21] and [44], that a single-valued, continuous, and likewise increasing, inverse function $x = \varphi(y)$ exists in the interval (A, B). Since it is increasing, if $\Delta x = 0$, $\Delta y = 0$,

and conversely; and due to continuity, $\Delta x \to 0$ implies $\Delta y \to 0$, and conversely. (The case of a decreasing function is exactly similar.)

THEOREM. *If $f(x)$ has a non-zero derivative $f'(x_0)$ at the point x_0, the inverse function $\varphi(y)$ has a derivative at the point $y_0 = f(x_0)$:*

$$\varphi'(y_0) = \frac{1}{f'(x_0)} . \tag{4}$$

Denoting corresponding increments of x and y by Δx and Δy, i.e.

$$\Delta x = \varphi(y_0 + \Delta y) - \varphi(y_0);$$

$$\Delta y = f(x_0 + \Delta x) - f(x_0) ,$$

and noting that both these differ from zero, we can write:

$$\frac{\Delta x}{\Delta y} = \frac{1}{\frac{\Delta y}{\Delta x}} .$$

As we have seen above, Δx and Δy tend simultaneously to zero, and the last equation leads to (4) in the limit. The present theorem can be put in the form of the following rule for differentiation of inverse functions: *the derivative of an inverse function is equal to unity divided by the derivative of the direct function at the corresponding point.*

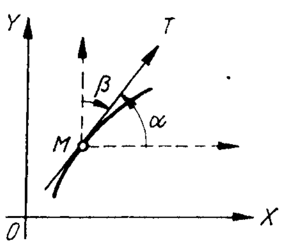

FIG. 53

The rule for differentiation of inverse functions has a simple geometrical interpretation [21]. The functions $x = \varphi(y)$ and $y = f(x)$ have the same graph in the XOY plane, the only difference being that the axis of the independent variable is OY, and not OX, for the function $x = \varphi(y)$. On drawing the tangent MT, and recalling the geometrical significance of the derivative, we get:

$$f'(x) = \tan(OX, MT) = \tan \alpha,$$

$$\varphi'(y) = \tan(OY, MT) = \tan \beta ,$$

angle β as well as α being reckoned positive, as in Fig. 53.

But evidently, $\beta = \frac{1}{2}\pi - \alpha$, and hence:

$$\tan \beta = \frac{1}{\tan \alpha} , \qquad \text{i.e.} \quad \varphi'(y) = \frac{1}{f'(x)} .$$

If $x = \varphi(y)$ is the inverse of $y = f(x)$, the converse is evidently true, i.e. $y = f(x)$ can be considered the inverse of $x = \varphi(y)$.

We use the rule for differentiation of inverse functions for the exponential function.

XIII. $y = a^x$ $(a > 0)$.

The inverse function in this case will be:

$$x = \varphi(y) = \log_a y\,,$$

and by VII:

$$\varphi'(y) = \frac{1}{y} \cdot \frac{1}{\log a}\,,$$

whence by the rule for differentiation of inverse functions:

$$y' = \frac{1}{\varphi'(y)} = y \log a, \text{ or } (a^x)' = a^x \log a.$$

In the particular case of $a = \mathrm{e}$, we have:

$$(\mathrm{e}^x)' = \mathrm{e}^x.$$

The formula obtained, together with the rule for differentiation of a function of a function, enables us to calculate the derivative of a power function.

XIV. $y = x^n$ ($x > 0$; n is any real number).

This function is defined, and is positive, for all $x > 0$ [19].

Using the definition of logarithm, we can express our function as a function of a function:

$$y = x^n = \mathrm{e}^{n \log x}$$

Using the rule for differentiation of a function of a function, we get:

$$y' = \mathrm{e}^{n \log x} \cdot \frac{n}{x} = x^n \cdot \frac{n}{x} = nx^{n-1}.$$

This result can be easily generalized for the case of negative x, provided the function exists for such values, as for instance $y = x^{1/3} = \sqrt[3]{x}$.

We use the rule for differentiation of inverse functions for obtaining the derivatives of the inverse circular functions.

XV. $y = \arcsin x$.

We consider the principal value of this function [24], i.e. the arc lying in the interval $(-\pi/2, +\pi/2)$. We can consider this as the inverse of function $x = \sin y$, and in accordance with the rule for differentiation of inverse functions, we have:

$$y'_x = \frac{1}{x'_y} = \frac{1}{\cos y} = \frac{1}{\sqrt{1 - \sin^2 y}} = \frac{1}{\sqrt{1 - x^2}}\,,$$

where the plus sign must be taken with the radical, since $\cos y$ has a positive sign in the interval $(-\pi/2, +\pi/2)$. We can similarly obtain:

$$(\text{arc} \ \cos x)' = -\frac{1}{\sqrt{1-x^2}} \, .$$

the principal value of arc cos x being taken, i.e. the arc contained in the interval $(0, \pi)$.

XVI. $y = \text{arc} \tan x$.

The principal value of arc $\tan x$ lies in the interval $(\pi/2, \pi/2)$, and we can consider this function as the inverse of $x = \tan y$; hence:

$$y'_x = \frac{1}{x'_y} = \frac{1}{\dfrac{1}{\cos^2 y}} = \cos^2 y = \frac{1}{1+\tan^2 y} = \frac{1}{1+x^2} \, .$$

We obtain similarly:

$$(\text{arc} \ \cot x)' = -\frac{1}{1+x^2} \, .$$

XVII. We also consider the differentiation of a function of the form:

$$y = u^v \, ,$$

where u and v are functions of x (exponential function).

We can write:

$$y = e^{v \log u},$$

and on using the rule for differentiation of a function of a function, we obtain:

$$y' = {}^{v \log u} \ (v \ \log \ u)' \, .$$

Using the rule for differentiation of a product, and differentiating $\log u$ as a function of a function of x, we finally have:

$$y' = e^{v \log u} \left(v' \ \log \ u + \frac{v}{u} u'\right)$$

or

$$y' = u^v \left(v' \ \log \ u + \frac{v}{u} u'\right) .$$

49. Table of derivatives, and examples. A list follows of all the rules that we have deduced for differentiation.

1. $(c)' = 0$.
2. $(cu)' = cu'$.
3. $(u_1 + u_2 + \ldots + u_n)' = u'_1 + u'_2 + \ldots + u'_n$.

4. $(u_1 u_2 \ldots u_n)' = u_1' u_2 u_3 \ldots u_n + u_1 u_2' u_3 \ldots u_n + \ldots +$
$+ u_1 u_2 u_3 \ldots u_n'$.

5. $\left(\frac{u}{v}\right)' = \frac{u' v - v' u}{v^2}$.

6. $(x^n)' = nx^{n-1}$ and $(x)' = 1$.

7. $(\log_a x)' = \frac{1}{x} \cdot \frac{1}{\log a}$ and $(\log x)' = 1/x$.

8. $(e^x)' = e^x$ and $(a^x)' = a^x \log a$.

9. $(\sin x)' = \cos x$.

10. $(\cos x)' = -\sin x$.

11. $(\tan x)' = \frac{1}{\cos^2 x}$.

12. $(\cot x)' = -\frac{1}{\sin^2 x}$

13. $(\arcsin x)' = \frac{1}{\sqrt{1 - x^2}}$.

14. $(\arccos x)' = -\frac{1}{\sqrt{1 - x^2}}$.

15. $(\arctan x)' = \frac{1}{1 + x^2}$.

16. $(\operatorname{arc\,cot} x)' = -\frac{1}{1 + x^2}$.

17. $(u^v)' = vu^{v-1} u' + u^v \log uv'$.

18. $y_x' = y_u' \cdot u_x'$ (y depends on x through the medium of u).

19. $x'_y = \frac{1}{y'_x}$.

We use the above rules to solve a few examples.

1. $y = x^3 - 3x^2 + 7x - 10$.

Using Rules 3, 6, and 2, we obtain $y' = 3x^2 - 6x + 7$.

2. $y = \frac{1}{\sqrt[3]{x^2}} = x^{-2/3}$.

Using Rule 6, we obtain: $y' = -\frac{2}{3} x^{-5/3} = -\frac{2}{3x\sqrt[3]{x^2}}$.

3. $y = \sin^2 x$.
We put $u = \sin x$ and use Rules 18, 6, and 9:

$$y' = 2u \cdot u' = 2 \sin x \cos x = \sin 2x.$$

4. $y = \sin (x^2)$.
We put $u = x^2$ and use the same rules:

$$y' = \cos u \cdot u' = 2x \cos (x^2).$$

5. $y = \log(x + \sqrt{1 + x^2})$,

We first put $u = x + \sqrt{x^2 + 1}$, then $v = x^2 + 1$, and use Rules 18 (twice), 7, 3, and 6:

$$y' = \frac{u'}{u} = \frac{(1 + v^{\frac{1}{2}})'}{u} = \frac{1}{u}\left(1 + \frac{1}{2}v^{-\frac{1}{2}}v'\right) = \frac{1 + xv^{-\frac{1}{2}}}{u} =$$

$$= \frac{1}{x + \sqrt{x^2 + 1}} \cdot \frac{x + \sqrt{x^2 + 1}}{\sqrt{x^2 + 1}} = \frac{1}{\sqrt{x^2 + 1}}.$$

6. $y = \left(\frac{x}{2x + 1}\right)^n$.

We put $u = \frac{x}{2x + 1}$ and use Rules 18, 6 and 5:

$$y' = nu^{n-1}u' = nu^{n-1}\frac{2x + 1 - 2x}{(2x + 1)^2} = \frac{nx^{n-1}}{(2x + 1)^{n+1}}.$$

7. $y = x^x$.

Using Rule 17, we obtain:

$$y' = x^{x-1} \cdot x + x^x \log x = x^x(1 + \log x).$$

8. The function y is given as an implicit function of x by the equation:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} - 1 = 0 \qquad (5)$$

The problem is to find the derivative of y.

If we solved the given equation for y, obtaining $y = f(x)$, the left-hand side of the equation would evidently be identical with zero after substituting $y = f(x)$. But the derivative of zero is the same as the derivative of a constant equal to zero, and hence we must obtain zero on differentiating the left-hand side of the given equation with respect to x, y being taken as the function of x given by this equation:

$$\frac{2x}{a^2} + \frac{2y}{b^2}y' = 0, \quad \text{whence} \quad y' = -\frac{b^2x}{a^2y}.$$

We see that in this case y' is expressed in terms of y as well as x; but we did not need to solve equation (5) for y, i.e. obtain an explicit expression for the function, in order to obtain the derivative.

Equation (5) represents an ellipse, as is known from analytic geometry, and the expression obtained for y' gives the slope of the tangent to the ellipse at the point with coordinates (x, y).

50. The concept of differential. Let Δx be the arbitrary increment of the independent variable, which we still take as not depending on x. We call it the *differential of the independent variable*, and denote it by the symbol Δx or $\mathrm{d}x$. The latter symbol in no circumstances stands for the product of d and x, being used only as a symbol for denoting an *arbitrary* quantity, *independent* of x, which we take as the increment of the independent variable.

The product of the derivative of a function and the differential of the independent variable is called the differential of the function.

The differential of a function is denoted by the symbol $\mathrm{d}y$ or $\mathrm{d}f(x)$:

$$\mathrm{d}y \text{ or } \mathrm{d}f(x) = f'(x)\mathrm{d}x. \tag{6}$$

This formula gives us an expression for the derivative in the form of the quotient of the differentials:

$$f'(x) = \frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}f(x)}{\mathrm{d}x}.$$

The differential of a function does not coincide with its increment. To explain the difference between these concepts, we turn to the graph of a function. We take a certain point $M(x, y)$ on the graph, with a second point N. We draw the tangent MT, the ordinates corresponding to M and N, and the line MP parallel to axis OX (Fig. 54). We have:

$$\overline{MP} = \overline{M_1N_1} = \Delta x \quad (\text{or } \mathrm{d}x),$$

$PN = \Delta y$ (increment of y),

$$\tan\angle PMQ = f'(x),$$

whence

$$\mathrm{d}y = f'(x)\mathrm{d}x = \overline{MP}\tan\angle PMQ = \overline{PQ}.$$

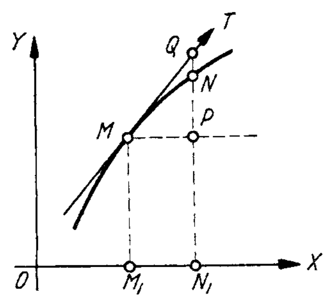

FIG. 54

The differential of the function, consisting of $\overline{PQ}$, does not coincide with $\overline{PN}$, which gives the increment of the function. The segment $\overline{PQ}$ gives the increment that would be obtained if we replaced segment $\overline{MN}$ of the curve in the interval $(x, x + \mathrm{d}x)$ by the segment $\overline{MQ}$ of the tangent, i.e. if we took the increment of the function as proportional to the increment of the independent variable in this interval, the coefficient of proportionality being taken as equal to the slope of the tangent MT, or, what amounts to the same thing, to the derivative $f'(x)$.

The segment $\overline{NQ}$ gives the difference between the differential and the increment. We show that, if Δx tends to zero, this difference is an infinitesimal of higher order with respect to Δx [36].

The ratio $\Delta y/\Delta x$ gives the derivative in the limit, hence [27]:

$$\frac{\Delta y}{\Delta x} = f'(x) + \varepsilon,$$

where ε is an infinitesimal along with Δx. We get from this equation:

$$\Delta y = f'(x)\Delta x + \varepsilon\Delta x$$

or

$$\Delta y = \mathrm{d}y + \varepsilon\Delta x ,$$

whence it is clear that the difference between $\mathrm{d}y$ and Δy is $(-\varepsilon\Delta x)$. But the ratio of $(-\varepsilon\Delta x)$ to Δx, equal to $(-\varepsilon)$, tends to zero along with Δx, i.e. the difference between $\mathrm{d}y$ and Δy is an infinitesimal of higher order with respect to Δx. We note that this difference can take either sign. Both Δx and the difference take the positive sign in our figure.

Formula (6) gives the rule for finding the differential of a function. We use it in some particular cases.

I. If c is a constant,

$$\mathrm{d}c = (c)'\mathrm{d}x = 0 \cdot \mathrm{d}x = 0 ,$$

i.e. *the differential of a constant is zero.*

II. $\mathrm{d}\,[cu(x)] = [cu(x)]'\mathrm{d}x = cu'(x)\mathrm{d}x = c\,\mathrm{d}u(x)$,

i.e. *a constant factor can be taken outside the differentiation sign.*

$$\text{III. } \mathrm{d}\,[u(x) + v(x) + w(x)] = [u(x) + v(x) + w(x)']\,\mathrm{d}x =$$

$$= [u'(x) + v'(x) + w'(x)]\mathrm{d}x = u'(x)\mathrm{d}x + v'(x)\mathrm{d}x + w'(x)\mathrm{d}x =$$

$$= \mathrm{d}u(x) + \mathrm{d}v(x) + \mathrm{d}w(x),$$

i.e. *the differential of a sum is equal to the sum of the differentials of the terms.*

$$\text{IV. } \mathrm{d}[u(x)v(x)w(x)] = [u(x)v(x)w(x)]'\mathrm{d}x =$$

$$= v(x)w(x)u'(x)\mathrm{d}x + u(x)w(x)v'(x)\mathrm{d}x + u(x)v(x)w'(x)\mathrm{d}x =$$

$$= v(x)w(x)\mathrm{d}u(x) + u(x)w(x)\mathrm{d}v(x) + u(x)v(x)\mathrm{d}w(x),$$

i.e. *the differential of a product is equal to the sum of the products of the differential of each with the remaining factors.*

We confine ourselves to the case of three factors. The same result is obtained for any finite number of factors.

$$\text{V.}\quad \mathrm{d}\,\frac{u(x)}{v(x)} = \left[\frac{u(x)}{v(x)}\right]'\mathrm{d}x = \frac{v(x)\,u'(x)\,\mathrm{d}x - u(x)\,v'(x)\,\mathrm{d}x}{[v(x)]^2} =$$

$$= \frac{v(x)\,\mathrm{d}u(x) - u(x)\,\mathrm{d}v(x)}{[v(x)]^2},$$

i.e. *the differential of a quotient (fraction) is equal to the denominator times the differential of the numerator, minus the numerator times the differential of the denominator, all divided by the square of the denominator.*

VI. We consider a function of a function, $y = f(u)$, where u is a function of x. We find $\mathrm{d}y$, assuming y dependent on x:

$$\mathrm{d}y = y'_x\,\mathrm{d}x = f'(u) \cdot u'_x\,\mathrm{d}x = f'(u)\mathrm{d}u\,,$$

i.e. *the differential of a function of a function has the same form as would be obtained by treating the auxiliary function as independent variable.*

We consider a numerical example, so as to compare the magnitudes of the increment of a function and its differential. We take the function:

$$y = f(x) = x^3 + 2x^2 + 4x + 10$$

and consider its increment:

$$f(2.01) - f(2) = 2.01^3 + 2\times 2.01^2 + 4\times 2.01 + 10 - \\ - (2^3 + 2\times 2^2 + 4\times 2 + 10).$$

On performing all the operations, we obtain the magnitude of the increment:

$$\Delta y = f(2.01) - f(2) = 0\cdot 240801\,.$$

Calculation of the differential is much easier. Here, $\mathrm{d}x = 2.10 - 2 = 0.01$, and the differential of the function is:

$$\mathrm{d}y = 3(x^2 + 4x + 4)\,\mathrm{d}x = (2\times 2^2 + 4.2 + 4)\times 0.01 = 0.24\,.$$

On comparing Δy and $\mathrm{d}y$, we see that they coincide to three decimal places.

51. Some differential equations. We have shown that replacing the increment of a function by its differential in the interval $(x, x + \mathrm{d}x)$ is equivalent to using the law of direct proportionality between the increment of the function and that of the independent variable, with the corresponding coefficient of proportionality. Furthermore, we know that this substitution involves an error that is an infinitesimal of higher order with respect to $\mathrm{d}x$. This is the basis of an application of the analysis of infinitesimals to the study of natural phenomena.

An effort may be made to split up a process into small elements when observing it, so that each element satisfies the law of direct proportionality on account of its smallness. An equation is obtained in the limit, represent-

ing the relationship between the independent variable, the function, and their differentials (or derivatives). This is called a *differential equation*, corresponding to the observed process. The task of finding the function itself from the differential equation is the task of *integrating the differential equation*.

When applying the analysis of infinitesimals to the study of some law of nature, a differential equation has to be derived for the law in question, then integrated. The latter task is usually far more difficult than the former, and more will be said about it later. We introduce the differential equations of a few very simple natural phenomena in the next sections.

1. *Barometric formula*. The atmospheric pressure p per unit area is evidently a function of the height above the earth's surface. We take a vertical cylindrical column of air of unit cross-sectional area. We consider two cross-sections, A and A_1, at heights h and $h + \mathrm{d}h$. On passing from A to A_1, the pressure p decreases (if $\mathrm{d}h < 0$) by an amount equal to the weight of air contained between A and A_1 in the cylinder. If $\mathrm{d}h$ is small, we can take the density ϱ of the air as approximately constant in this part of the cylinder. Since the base-area of the small cylinder AA_1 is unity, and its height $\mathrm{d}h$, its volume is $\mathrm{d}h$, and the required weight is $\varrho \mathrm{d}h$. Thus the decrease in p (for $\mathrm{d}h < 0$) is equal to $\varrho \mathrm{d}h$:

$$\mathrm{d}p = -\varrho \mathrm{d}h .$$

By Boyle's law, the density ϱ is proportional to the pressure p:

$$\varrho = cp \quad (c \text{ is a constant}) ,$$

and we finally obtain the differential equation:

$$\mathrm{d}p = -cp\,\mathrm{d}h \quad \text{or} \quad \frac{\mathrm{d}p}{\mathrm{d}h} = -cp .$$

2. *Chemical reaction of the first order*. Let a certain substance, of mass a, take part in a chemical reaction. We let x denote the part of this mass that has already taken part in the reaction at the instant t, measured from the start of the reaction. Evidently, x is a function of t. It can be taken approximately, for certain reactions, that the amount of substance $\mathrm{d}x$, taking part in the reaction in the interval of time from t to $t + \mathrm{d}t$, is proportional to $\mathrm{d}t$ for small $\mathrm{d}t$, as also to the amount of substance that has still not taken part in the reaction at the instant t:

$$\mathrm{d}x = c\,(a - x)\,\mathrm{d}t \quad \text{or} \quad \frac{\mathrm{d}x}{\mathrm{d}t} = c\,(a - x) .$$

We transform this differential equation, taking a new function $y = a - x$ instead of x, where y denotes the mass that has still not taken part in the reaction by the instant t. Noting that a is a constant, we have:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -\frac{\mathrm{d}x}{\mathrm{d}t} ,$$

so that the differential equation of a chemical reaction of the first order can be rewritten in the form:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -cy .$$

3. *Law of cooling.* We suppose that a certain body, heated to high temperature, is placed in a medium with a constant temperature of 0°. The temperature θ of the body on cooling will be a function of time t, which we measure from the instant of placing the body in the medium. The amount of heat dQ given up by the body in the interval dt can be taken as approximately proportional to the length dt of the interval and to the difference in temperature of the body and the medium at the instant t (Newton's law of cooling). We can then write:

$$dQ = c_1 \theta \, dt \quad (c_1 \text{ is a constant}).$$

Denoting the specific heat of the body by k, we have:

$$dQ = -k d\theta ,$$

where the minus sign is taken because $d\theta$ is negative in the present case (the temperature falls). On comparing these two expressions for dQ, we get

$$d\theta = -c\theta \, dt \left(c = \frac{c_1}{k}\right) \quad \text{or} \quad \frac{d\theta}{dt} = -c\theta ;$$

c is a constant, if we take the specific heat k as constant. All the differential equations that we have deduced above have the same form. They all express the proportionality of the derivative to the function itself, with a negative coefficient $(-c)$.

We showed in [38] that continuous interest on a basic capital a over t years yields an accumulated capital of ae^{kt}, where k is the percentage interest expressed as a fraction; we write:

$$y = ae^{kt}. \tag{7}$$

We obtain the derivative:

$$y' = ake^{kt} = ky , \tag{8}$$

i.e. we obtain here the same property of proportionality of the derivative to the function itself, which is why this property is referred to as the *law of compound interest.* We shall show later that function (7) gives *all* the solutions of differential equation (8) with *any* value of the constant a, in place of which we shall write C.

Thus, the solutions of our equations can be set in the form (replacing k by $-c$):

$$p(h) = Ce^{-ch}, \; y(t) = Ce^{-ct}, \; \theta(t) = Ce^{-ct} , \tag{9}$$

where C is a constant. We now indicate the physical significance of constant C in each of the above formulae. Putting $h = 0$ in the first formula, we get:

$$C = p(0) = p_0 ,$$

where p_0 is the atmospheric pressure at $h = 0$, i.e. at the earth's surface. The second formula gives us at $t = 0$:

$$C = y(0) ,$$

i.e. C is the mass that has not taken part in the reaction at the initial instant, which was earlier denoted by a. Finally, putting $t = 0$ in the third of the formulae (9), we see similarly that C is the initial temperature θ_0 of the body at the moment of placing it in the medium. We thus have these results:

$$p(h) = p_0 e^{-ch}, \ y(t) = ae^{-ct}, \ \theta(t) = \theta_0 e^{-ct}. \tag{10}$$

52. Estimation of errors. When a magnitude x is found in practice or is roughly calculated, an error Δx is obtained, called the *absolute error* of the observation or calculation. It does not characterize the accuracy of the observation. For instance, an error of about 1 cm in giving the length of a room would be permissible in practice, whereas the same error in determining the distance between two nearby objects (say the source and screen of a photometer) would point to great inaccuracy in the measurement. Hence follows the further concept of *relative error*, this being equal to the absolute value of the ratio $|\Delta x/x|$ of the absolute error to the measured magnitude.

We now suppose that a certain magnitude y is defined by the equation $y = f(x)$. An error Δx in defining x leads to a corresponding error Δy. For small values of Δx, Δy can be approximately replaced by the differential dy, so that the relative error in defining magnitude y is given by

$$\left|\frac{dy}{y}\right|.$$

Examples. **1.** The current i is given with a tangent galvanometer by the well-known formula:

$$i = c \tan \varphi.$$

Let $d\varphi$ be the error in reading angle φ:

$$di = \frac{c}{\cos^2 \varphi} d\varphi, \ \frac{di}{i} = \frac{c}{\cos^2 \varphi \cdot c \tan \varphi} d\varphi = \frac{2}{\sin 2\varphi} d\varphi,$$

whence it is clear that the relative error $|di/i|$ in determining i will be smaller, the nearer φ is to $45°$.

2. We take the product uv:

$$d(uv) = v\,du + u\,dv, \ \frac{d(uv)}{uv} = \frac{du}{u} + \frac{dv}{v},$$

whence it follows:

$$\left|\frac{d(uv)}{uv}\right| \leqslant \left|\frac{du}{u}\right| + \left|\frac{dv}{v}\right|,$$

i.e. *the relative error of a product is not greater than the sum of the relative errors in the terms.*

The same rule is obtained for a quotient, since :

$$d\frac{u}{v} = \frac{v\,du - u\,dv}{v^2}, \ \frac{d\frac{u}{v}}{\frac{u}{v}} = \frac{du}{u} - \frac{dv}{v};$$

$$\left|\frac{du\,v}{uv}\right| \leqslant \left|\frac{du}{u}\right| + \left|\frac{dv}{v}\right|.$$

3. We consider the formula for the area of a circle:

$$Q = \pi r^2, \ \mathrm{d}Q = 2\pi r\,\mathrm{d}r, \frac{\mathrm{d}Q}{Q} = \frac{2\pi r\,\mathrm{d}r}{\pi r^2} = 2\frac{\mathrm{d}r}{r},$$

i.e. the relative error in determining the area of a circle is by the above formula equal to twice the relative error in determining the radius.

4. We suppose that angle φ is defined by the logarithms of its sine and tangent. We have by the rules of differentiation:

$$\mathrm{d}(\log_{10}\sin\varphi) = \frac{\cos\varphi\,\mathrm{d}\varphi}{\log 10\times\sin\varphi}, \ \mathrm{d}(\log_{10}\tan\varphi) = \frac{\mathrm{d}\varphi}{\log 10\times\tan\varphi\times\cos^2\varphi}$$

whence

$$\mathrm{d}\varphi = \frac{\log 10\times\sin\varphi}{\cos\varphi}\,\mathrm{d}(\log_{10}\sin\varphi), \ \mathrm{d}\varphi = \log 10\times\sin\varphi\cos\varphi\,\mathrm{d}(\log_{10}\tan\varphi). \quad (11)$$

We suppose that we make the same error in finding $\log_{10}\sin\varphi$ and $\log_{10}\tan\varphi$ (this error depends on the number of decimal places in the logarithmic tables that we use). The first of formulae (11) gives a value of $\mathrm{d}\varphi$ of greater absolute value than that obtained with the second of formulae (11), since the product $\log 10\times\sin\varphi$ is divided in the first case, and multiplied in the second case, by $\cos\varphi$, and $|\cos\varphi| < 1$. It is thus better to calculate the angles using the table for $\log_{10}\tan\varphi$.

§ 4. Derivatives and differentials of higher orders

53. Derivatives of higher orders. The derivative $f'(x)$ of the function $y = f(x)$ is also a function of x, as we know. On differentiating it, we obtain a new function, called the *second derivative* or the *derivative of the second order* of the original function $f(x)$, and denoted by:

$$y'' \quad \text{or} \quad f''(x).$$

On differentiating the second derivative, we obtain the derivative of the third order, or simply, the third derivative:

$$y''' \quad \text{or} \quad f'''(x).$$

Using the operation of differentiation in this way, we obtain the derivative of any order n, $y^{(n)}$ or $f^{(n)}(x)$.

We consider some examples.

1. $y = \mathrm{e}^{ax}$, $y' = a\mathrm{e}^{ax}$, $y'' = a^2\mathrm{e}^{ax}, \ldots, y^{(n)} = a^n\mathrm{e}^{ax}$.

2. $y = (ax+b)^k$, $y' = ak(ax+b)^{k-1}$,

$$y'' = a^2 k(k-1)(ax+b)^{k-2}, \ldots,$$

$$y^{(n)} = a^n k(k-1)(k-2)\ldots(k-n+1)(ax+b)^{k-n}.$$

3. We know that:

$$(\sin x)' = \cos x = \sin\left(x + \frac{1}{2}\pi\right), \ (\cos x)' = -\sin x = \cos\left(x + \frac{1}{2}\pi\right),$$

i.e. differentiation of $\sin x$ and $\cos x$ amounts to increasing the argument by $\pi/2$, and hence:

$$(\sin x)'' = \left[\sin\left(x + \frac{1}{2}\pi\right)\right]' = \sin\left(x + 2\cdot\frac{1}{2}\pi\right)\cdot\left(x + \frac{1}{2}\pi\right)' = \\ = \sin\left(x + 2\cdot\frac{1}{2}\pi\right),$$

and in general:

$$(\sin x)^{(n)} = \sin\left(x + n\cdot\frac{1}{2}\pi\right) \quad \text{and} \quad (\cos x)^{(n)} = \cos\left(x + n\cdot\frac{1}{2}\pi\right).$$

4. $y = \log(1 + x), y' = \dfrac{1}{1+x}, \ y'' = -\dfrac{1}{(1+x)^2},$

$$y''' = \frac{1\cdot 2}{(1+x)^3} \ldots, \ y^{(n)} = (-1)^{n+1}\frac{(n-1)!}{(1+x)^n}.$$

5. We consider the sum of functions:

$$y = u + v + w.$$

Using the rule for differentiation of a sum, and assuming that the corresponding derivatives of functions u, v and w exist, we have:

$$y' = u' + v' + w', \quad y'' = u'' + v'' + w'', \quad y^{(n)} = u^{(n)} + v^{(n)} + w^{(n)},$$

i.e. *the derivative of any order of a sum is equal to the sum of the derivatives of the same order.* For example:

$y = x^3 - 4x^2 + 7x + 10$; $y' = 3x^2 - 8x + 7$; $y'' = 6x - 8$; $y''' = 6$; $y^{(4)} = 0$ and generally, $y^{(n)} = 0$ for $n > 3$.

It can be shown in the same way that, in general, the *n-th derivative of a polynomial of degree m is zero if $n > m$.*

We now consider the derivatives of the product of two functions u, v viz., $y = uv$. We use the rules for differentiation of a product and a sum:

$$y' = u'v + uv', \ y'' = u''v + u'v' + u'v' + uv'' = \\ = u''v + 2u'v' + uv'', \quad y''' = u'''v + u''v' + 2u''v' + 2u'v'' + \\ + u'v'' + uv''' = u'''v + 3u''v' + 3u'v'' + uv'''.$$

We note the following rule: *in order to obtain the n-th derivative of the product uv, $(u + v)^n$ should be determined by Newton's binomial*

formula, and the exponents of powers of u and v in the result should be taken to indicate the orders of derivatives, the zero powers ($u^0 = v^0 = 1$) appearing in the extreme terms of the result being taken as the functions themselves.

This rule is known as Leibniz's theorem, and is written symbolically as:

$$y^{(n)} = (u + v)^{(n)}.$$

We prove this theorem by induction. We suppose that the rule is true for the nth derivative, i.e:

$$y^{(n)} = (u + v)^{(n)} = u^{(n)} v + \frac{n}{1} u^{(n-1)} v' + \frac{n(n-1)}{2!} u^{(n-2)} v'' + \ldots +$$

$$+ \frac{n(n-1)\ldots(n-k+1)}{k!} u^{(n-k)} v^{(k)} + \ldots + uv^{(n)}. \tag{1}$$

To obtain $y^{(n+1)}$, we differentiate the above sum with respect to x. The general term $u^{(n-k)} v^{(k)}$ of the sum gives, by the rule for differentiation of a product, the derivative $u^{(n-k+1)} v^{(k)} + u^{(n-k)} v^{(k+1)}$. But this latter sum can be written symbolically as:

$$u^{n-k} v^k (u + v).$$

In fact, we have regarded the orders of the derivatives as indices and we have factorized the expression

$$u^{(n-k+1)} v^{(k)} + u^{(n-k)} v^{(k+1)}$$

Hence we see that $y^{(n+1)}$ must be obtained by multiplying each term of the sum (1), and hence all the sum, symbolically by $(u + v)$, so that:

$$y^{(n+1)} = (u + v)^{(n)} \times (u + v) = (u + v)^{(n+1)}.$$

We have shown that if Leibniz's theorem is true for any given n, it is true for $(n + 1)$. But we have shown directly that it is true for $n = 1, 2, 3$; and hence it is true for all n.

We take as an example:

$$y = e^x (3x^2 - 1)$$

and find $y^{(100)}$:

$$y^{(100)} = (e^x)^{(100)} (3x^2 - 1) + \frac{100}{1} (e^x)^{(99)} (3x^2 - 1)' +$$

$$+ \frac{100 \cdot 99}{1 \cdot 2} (e^x)^{(98)} (3x^2 - 1)'' + \ldots + e^x (3x^2 - 1)^{(100)}.$$

All the derivatives of a polynomial of the second degree are zero as from the third order, and $(e^x)^{(n)} = e^x$, whence:

$$y^{(100)} = e^x (3x^2 - 1) + 100\, e^x \times 6x + 4950\, e^x \times 6 =$$
$$= e^x (3x^2 + 600x + 29{,}699).$$

54. Mechanical significance of the second derivative. We consider the motion of a point on a straight line:

$$s = f(t)\,,$$

where, as usual, t denotes time, and s is the path, measured from some fixed point of the line. We obtain the *velocity* of the motion by differentiating once with respect to t:

$$v = f'(t)\,.$$

We obtain the second derivative as the limit of the ratio $\Delta v/\Delta t$ as Δt tends to zero. This ratio characterizes the rate of change of velocity in the interval Δt, and gives the average acceleration in this interval, whilst the limit of the ratio as $\Delta t \to 0$ gives the *acceleration* w of the observed motion at time t:

$$w = f''(t)\,.$$

We take $f(t)$ as a polynomial of the second degree:

$$s = at^2 + bt + c, \quad v = 2at + b, \quad w = 2a,$$

i.e. the acceleration w is constant, and the coefficient $a = w/2$. etting $t = 0$, we get $b = v_0$, i.e. the coefficient b is equal to the initial velocity, and $c = s_0$, i.e. c is equal to the distance of the point at $t = 0$ from the origin on the line. On putting these values for a, b, and c in the expression for s, we get the formula for the path with uniformly accelerated $(w > 0)$ or uniformly decelerated $(w < 0)$ motion:

$$s = \frac{1}{2} wt^2 + v_0 t + s_0.$$

In general, on knowing the law for the change of path, we can find the acceleration w by differentiating twice with respect to t, and hence find the force f producing the motion, since, by Newton's second law, $f = mw$, where m is the mass of the moving point.

All the above applies only to linear motion. In the case of curvilinear motion, as is shown in mechanics, $f''(t)$ gives only the projection of the acceleration vector on the tangent to the trajectory.

We take the example of a point M oscillating harmonically on a line, so that its distance s from a fixed point O of the line is defined by the formula:

$$s = a \sin\left(\frac{2\pi}{\tau} t + \omega\right),$$

where the amplitude a, the period of oscillation τ, and the phase ω are constants. Differentiations give us the velocity v and the force f:

$$v = \frac{2\pi a}{\tau} \cos\left(\frac{2\pi}{\tau} t + \omega\right), \quad f = mw =$$

$$-\frac{4\pi^2 m}{\tau^2} a \sin\left(\frac{2\pi}{\tau} t + \omega\right) = -\frac{4\pi^2 m}{\tau^2} s,$$

i.e. the force is proportional in magnitude to the length of the interval $\overline{OM}$ and acts in the opposite direction. In other words, the force is always directed from the point M to the point O, being proportional to their distance apart.

55. Differentials of higher orders. We now introduce the concept of higher order differentials of a function $y = f(x)$. Its differential

$$\mathrm{d}y = f'(x)\mathrm{d}x$$

is clearly a function of x, though it must be remembered that the differential $\mathrm{d}x$ of the independent variable is reckoned as independent of x [50], so that it must be taken outside the differentiation sign as a constant factor on further differentiation. The differential of $\mathrm{d}y$ can be obtained by treating it as a function of x; this is called the second order differential of the original function $f(x)$ and is denoted by d^2y or $\mathrm{d}^2f(x)$:

$$\mathrm{d}^2 y = \mathrm{d}(\mathrm{d}y) = [f'(x)\mathrm{d}x]' \,\mathrm{d}x = f''(x)\mathrm{d}x^2.$$

On obtaining the differential of this further function of x, we arrive at the third order differential:

$$\mathrm{d}^3 y = \mathrm{d}(\mathrm{d}^2 y) = [f''(x)\mathrm{d}x^2]' \,\mathrm{d}x = f'''(x)\mathrm{d}x^3 ,$$

and in general, we arrive by successive differentiation at the concep: of the nth order differential of function $f(x)$, which is expressed ast

$$\mathrm{d}^n f(x) \quad \text{or} \quad \mathrm{d}^n y = f^{(n)}(x)\mathrm{d}x^n. \tag{2}$$

This formula allows of the expression of the nth derivative as a fraction:

$$f^{(n)}(x) = \frac{\mathrm{d}^n y}{\mathrm{d}x^n} . \tag{3}$$

We now consider a function of a function, $y = f(u)$, where u is a function of some independent variable. We know [50] that the first differential of this function has the same form as when u is the independent variable:

$$dy = f'(u)du.$$

Formula (2) is no longer valid for obtaining higher order differentials, since we are not justified in treating du as a constant when u is not the independent variable. For example, we obtain the second differential by using the rule for finding the differential of a product, with the result:

$$d^2 y = d[f'(u)du] = du\, d[f'(u)] + f'(u)\, d(du) = f''(u)du^2 + f'(u)d^2u,$$

which has the extra term $f'(u)d^2 u$, compared with formula (2).

If u is the independent variable, du must be treated as constant and $d^2 u = 0$. We now take u as a linear function of the independent variable t, i.e.,

$$u = at + b.$$

Now, $du = a\, dt$, i.e., du is again a constant, so that the higher order differentials of the function of a function are given by (2):

$$d^n f(u) = f^{(n)}(u)\, du^n,$$

i.e. *formula* (2) *for the higher order differentials is valid when x is either the independent variable, or a linear function of the independent variable.*

56. Finite differences of functions. We denote the increment of the independent variable by h. The corresponding increment of $y = f(x)$ will be:

$$\Delta y = f(x + h) - f(x). \tag{4}$$

An alternative name is *the first-order difference of* function $f(x)$. This difference is, for its part, also a function of x, and we can find its difference by substracting from its value at $(x + h)$ its value at x. This new difference is called the *second-order difference* of the original function $f(x)$ and is denoted by $\Delta^2 y$. We can easily express $\Delta^2 y$ in terms of values of $f(x)$:

$$\Delta^2 y = [f(x + 2h) - f(x + h)] - [f(x + h) - f(x)] =$$
$$= f(x + 2h) - 2f(x + h) + f(x). \tag{5}$$

This second order difference is also a function of x, and the difference of this function can be defined, giving the *third-order difference* of the original function $f(x)$, denoted by $\Delta^3 y$. Replacing x by $(x + h)$ on the right of (5), and subtracting the right of (5) from the result, we obtain the expression for $\Delta^3 y$:

$$\begin{aligned}\Delta^3 y = [f(x + 3h) - 2f(x + 2h) + f(x + h)] - \\ - [f(x + 2h) - 2f(x + h) + f(x)] = \\ = f(x + 3h) - 3f(x + 2h) + 3f(x + h) - f(x).\end{aligned}$$

We can thus go on to define the difference of any order, the nth order difference $\Delta^n y$ being expressed as follows in terms of values of $f(x)$:

$$\begin{aligned}\Delta^n y = f(x + nh) - \frac{n}{1!} f(x + \overline{n + 1}h) + \\ + \frac{n(n-1)}{2!} f(x + \overline{n - 2}h) - \ldots + \qquad (6) \\ + (-1)^k \frac{n(n-1)\ldots(n-k+1)}{k!} f(x + \overline{n - k}h) + \ldots + (-1)^n f(x).\end{aligned}$$

We have seen above that this formula is true for $n = 1$, 2, and 3. A rigorous proof requires us to pass from n to $(n + 1)$ in the usual way. We note that $(n + 1)$ values of $f(x)$, for the values of the argument: $x, x + h, x + 2h, \ldots, x + nh$, are needed for calculating $\Delta^n y$. These values of the argument form an arithmetical progression with difference h, or as we say, represent equidistant values.

For small h, Δy differs little from the differential dy. Similarly, higher-order differences give approximate values of the differentials of corresponding order, and conversely. Whilst we lack an analytic expression for a function, we may be given a table for it for equidistant values of the argument; we cannot then calculate the various derivatives of the function accurately, but we can obtain approximate values by calculating the ratio $\Delta^n y/\Delta x^n$, instead of using the accurate formula (3). As an example, we give a table of differences and differentials of the function $y = x^3$ in the interval (2, 3), taking:

$$\Delta x = h = 0.1.$$

In setting up this table, the successive values of $y = x^3$ were calculated, then the values of Δy obtained from these, by subtraction in accordance with formula (4), then the values of $\Delta^2 y$ obtained from the values of Δy by further subtractions, and so on. This method of calculating the differences successively is naturally simpler than

using formula (6). The differentials were calculated in the ordinary way, the formulae being given at the top of the table, where we must take $dx = h = 0.1$.

We compare the accurate and approximate values of the second derivative y'' for $x = 2$. Here, $y'' = 6x$ and $y'' = 12$ with $x = 2$. The approximate value is given by the ratio $\Delta^2 y/h^2$, and we have for $x = 2$:

$$\frac{0.126}{(0.1)^2} = 12.6.$$

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$	$dy = 3x^2dx$	$d^2y = 6x dx^2$	$d^3y = 6dx^3$	$d^4y = 0$
2	8.000	1.261	0.126	0.006	0	1.200	0.120	0.006	0
2.1	9.261	1.387	0.132	0.006	0	1.323	0.126	0.006	0
2.2	10.648	1.519	0.138	0.006	0	1.452	0.132	0.006	0
2.3	12.167	1.657	0.144	0.006	0	1.587	0.138	0.006	0
2.4	13.824	1.801	0.150	0.006	0	1.728	0.144	0.006	0
2.5	15.625	1.951	0.156	0.006	0	1.875	0.150	0.006	0
2.6	17.576	2.107	0.162	0.006	0	2.028	0.156	0.006	0
2.7	19.683	2.269	0.168	0.006	—	2.187	0.162	0.006	—
2.8	21.952	2.437	0.174	—	—	2.352	0.168	—	—
2.9	24.389	2.611	—	—	—	2.523	—	—	—
3	27.000	—	—	—	—	—	—	—	—

If $f(x)$ is a polynomial of x:

$$y = f(x) = a_0 x^m + a_1 x^{m-1} + a_2 x^{m-2} + \dots + a_{m-1} x + a_m,$$

it is easily seen that, on calculating Δy by (4), we get a polynomial of degree $(m - 1)$ for Δy, with highest term $ma_0 hx^{m-1}$. Thus, with $y = x^3$, Δy is a polynomial of second degree in x, $\Delta^2 y$ is a polynomial of first degree, $\Delta^3 y$ is constant, and $\Delta^4 y$ is zero (see table). It is suggested as an exercise to the reader to show that the values of $d^2 y$ must be one step behind those of $\Delta^2 y$ in our example, as is obvious from the table.

§ 5. Application of derivatives to the study of functions.

57. Tests for increasing and decreasing functions. Knowledge of the derivative enables us to study the various properties of a function. We begin with the simplest and most basic question, of whether a function is increasing or decreasing.

The function $f(x)$ is said to be increasing in an interval if it increases correspondingly with the variable thraughout that interval, i.e. *if*

$$f(x+h) - f(x) > 0 \quad \text{for} \quad h > 0.$$

On the other hand, if

$$f(x+h) - f(x) < 0 \quad \text{for} \quad h > 0,$$

the function is said to be decreasing.

Turning to the graph of the function, the interval in which the function is increasing corresponds to the section of the graph for which greater values of x imply greater values of y. If OX is directed to the right, and OY upwards, as in Fig. 55, the interval of increase of the function will correspond to the part of the graph where movement along it to the right in the direction of increasing abscissa implies movement upwards. An interval of decrease, on the other hand, corresponds to a part of the curve where we move downwards on moving along the curve to the right. In Fig. 55, the interval of increase corresponds to section AB of the graph, and the interval of decrease to BC. It is immediately clear from the figure that in the first section, the tangent forms with the direction of OX an angle α measured from OX to the tangent, the tangent of which is positive but the tangent of this angle is also the first derivative $f'(x)$. In section BC, on the contrary, the direction of the tangent forms with the direction of OX an angle α (in the fourth quadrant), the tangent of which is negative, i.e. $f'(x)$ is negative in this case. Putting these results together, we arrive at the following rule: *a function is increasing in the intervals for which $f'(x)$ is positive, and is decreasing in intervals where $f'(x) < 0$.*

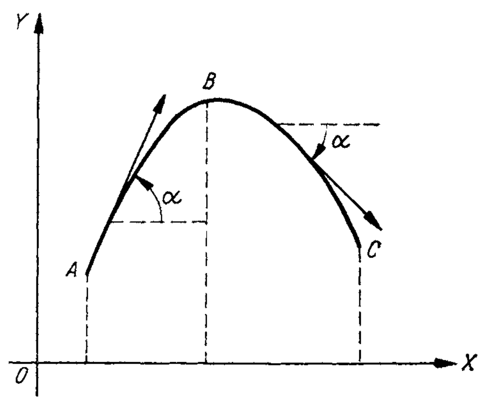

FIG. 55

We have arrived at this rule by using the figure. A rigorous analytic proof is given later. We now make use of the rule in some examples.

1. We prove the inequality:

$$\sin x > x - \frac{x^3}{6} \quad \text{for } x > 0.$$

For this, we form the difference:

$$f(x) = \sin x - \left(x - \frac{x^3}{6}\right).$$

We find the derivative $f'(x)$:

$$f'(x) = \cos x - 1 + \frac{x^2}{2} = \frac{x^2}{2} - (1 - \cos x) =$$

$$= \frac{x^2}{2} - 2\sin^2\frac{x}{2} = 2\left[\left(\frac{1}{2}x\right)^2 - \left(\sin\frac{1}{2}x\right)^2\right].$$

Since the absolute value of the arc is greater than its sine, we can say that $f'(x) > 0$ in the interval $(0, \infty)$, i.e. $f(x)$ is increasing in this interval; and $f(0) = 0$, so that:

$$f(x) = \sin x - \left(x - \frac{x^3}{6}\right) > 0 \quad \text{for } x > 0,$$

i.e.

$$\sin x > x - \frac{x^3}{6} \quad \text{for } x > 0.$$

2. It can be shown in exactly the same way that:

$$x > \log(1 + x) \text{ for } x > 0.$$

We form the difference:

$$f(x) = x - \log(1 + x),$$

whence

$$f'(x) = 1 - \frac{1}{1 + x}.$$

It is evident from this expression that $f'(x) > 0$ for $x > 0$, i.e. $f(x)$ is increasing in the interval $(0, +\infty)$; but $f(0) = 0$, so that:

$$f(x) = x - \log(1 + x) > 0 \text{ for } x > 0,$$

i.e.

$$x > \log(1 + x) \text{ for } x > 0.$$

3. We take Kepler's equation, discussed in [31]:

$$x = q\sin x + a \quad (0 < q < 1).$$

We can write this in the form:

$$f(x) = x - q\sin x - a = 0.$$

We get for the derivative:

$$f'(x) = 1 - q\cos x.$$

We note that the absolute value of $q \cos x$ is less than unity, since q lies between zero and unity by hypothesis; we can thus say that $f'(x) > 0$ for any x, so that $f(x)$ is increasing in the interval $(-\infty, +\infty)$ and can there-

fore be zero only once, i.e. Kepler's equation cannot have more than one real root.

If the constant a is a multiple of π, i.e. $a = k\pi$, where k is an integer, we obtain $f(k\pi) = 0$ on substituting $x = k\pi$ directly, so that $x = k\pi$ is the unique root of Kepler's equation. If a is not a multiple of π, we can find an integer k such that

$$k\pi < a < (k + 1)\,\pi .$$

Substituting $x = k\pi$ and $(k + 1)\,\pi$, we get:

$$f(k\pi) = k\pi - a < 0\,,$$
$$f(\overline{k + 1}\pi) = (k + 1)\,\pi - a > 0 .$$

But if $f(k\pi)$ and $f(\overline{k + 1}\pi)$ have different signs, $f(x)$ must be zero in the interval $(k\pi, \overline{k + 1}\pi)$ [35], i.e. the single root of Kepler's equation lies in this interval.

4. We take the equation:

$$f(x) = 3x^5 - 25x^3 + 60x + 15 = 0 .$$

We find the derivative $f'(x)$ and set it equal to zero:

$$f'(x) = 15x^4 - 75x^2 + 60 = 15(x^4 - 5x^2 + 4) = 0 .$$

We solve this biquadratic equation and find that $f'(x)$ is zero for

$$x = -2,\ -1,\ +1,\ \text{and}\ +2 .$$

We can now divide the total interval $(-\infty, +\infty)$ into five intervals:

$$(-\infty, -2),\ (-2, -1),\ (-1, 1),\ (1, 2),\ (2, +\infty)\,,$$

so that $f'(x)$ has the same sign in each, $f(x)$ being therefore monotonic, i.e. either increasing or decreasing, and hence having not more than one root in each interval. If $f(x)$ has different signs at the ends of a given interval, $f(x) = 0$ has one root in the interval; whereas if it has the same sign, there is no root in the interval. Thus, by finding the signs of $f(x)$ at the ends of the five intervals above, we can find the number of roots of the equation.

To find the signs of $f(x)$ at $x = \pm\infty$, we write $f(x)$ in the form:

$$f(x) = x^5\left(3 - \frac{25}{x^2} + \frac{60}{x^4} + \frac{15}{x^5}\right).$$

For $x \to -\infty$, $x^5 \to -\infty$, and the expression in brackets tends to 3, so $f(x) \to -\infty$. Similarly it can be seen that $f(x) \to +\infty$ for $x \to +\infty$. On substituting the values $x = -2, -1, 1$, and 2, we get the following table:

x	$-\infty$	-2	-1	1	2	$+\infty$
$f(x)$	$-$	$-$	$-$	$+$	$+$	$+$

We see that $f(x)$ has different signs only at the ends of the interval $(-1, 1)$, so that the equation concerned has only one real root, which lies in this interval.

We defined above a function increasing or decreasing in an interval. A function is sometimes said to be increasing or decreasing at a point $x = x_0$. The exact meaning is: $f(x)$ is increasing at $x = x_0$, if $f(x) < < f(x_0)$ for $x < x_0$, and $f(x) > f(x_0)$ for $x > x_0$, x being taken sufficiently close to x_0. Similarly for a function decreasing at a point. The concept of derivative leads directly to a sufficient condition for a function to be increasing or decreasing at a point, viz., if $f'(x_0) > 0$, $f(x)$ is increasing at x_0, and if $f'(x_0) < 0$, $f(x)$ is decreasing at x_0. If, e.g., $f'(x_0) > 0$, the ratio:

$$\frac{f(x_0 + h) - f(x_0)}{h},$$

with limit $f'(x_0)$, will also be positive for all h with sufficiently smal absolute values, i.e. numerator and denominator will have the same sign. In other words, we shall have $f(x_0 + h) - f(x_0) > 0$ for $h > 0$, and $f(x_0 + h) - f(x_0) < 0$ for $h < 0$, so that $f(x)$ is increasing at x_0.

58. Maxima and minima of functions. We again turn to the graph of some function $f(x)$ (Fig. 56). We have a successive alternation of intervals of increase and decrease in this case. Segment AM_1 corresponds to an interval of increase, the next segment $M_1 M_2$ to one of decrease, the next, $M_2 M_3$ again to one of increase, and so on. Peaks of the curve separate intervals of increase from those of decrease. Take the peak M_1, for example; its ordinate is greater than all the ordinates of the curve that are sufficiently close to the peak, whether to the left or right. Such a peak is said to correspond to a maximum of $f(x)$.

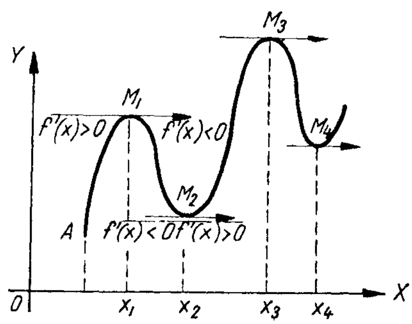

FIG. 56

This leads us to the following general analytic definition: *the function $f(x)$ attains a maximum at the point $x = x_1$ if its value $f(x_1)$ at this point is greater than its values at all neighbouring points,* i.e. *if the increment of the function*

$$f(x_1 + h) - f(x_1) < 0$$

for all positive or negative h, sufficiently small in absolute value.

We now consider the peak M_2. The ordinate in this case is less than all neighbouring ordinates, whether on the left or right; the peak is said to correspond to a minimum of the function, which is defined analytically as follows: *the function $f(x)$ attains a minimum at the point $x = x_2$, if we have*

$$f(x_2 + h) - f(x_2) > 0$$

for all positive or negative h, sufficiently small in absolute value.

We see from the figure that the tangents at peaks corresponding to either maxima or minima of the function $f(x)$ lie parallel to the axis OX, i.e. their slope $f'(x)$ is zero. The tangent to a curve can be parallel to OX, however, elsewhere than at a peak. In Fig. 57, for example, point M of the curve is not a peak, yet the tangent at M is parallel to OX.

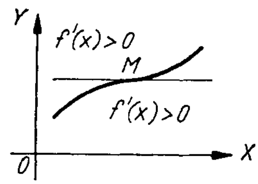

FIG. 57

Suppose that $f'(x)$ becomes zero for some value $x = x_0$, i.e. the tangent at the corresponding point of the graph is parallel to OX. We consider the sign of $f'(x)$ for x near x_0. We take the following three cases:

I. For x less than, and sufficiently near x_0, $f'(x)$ is positive, whilst $f'(x)$ is negative for x greater than, and sufficiently near x_0, i.e. in other words, as x passes through x_0, $f'(x)$ passes through zero from positive to negative values.

In this case, we have an interval of increase to the left of $x = x_0$, and an interval of decrease to the right, i.e. $x = x_0$ corresponds to a peak of the curve, where $f(x)$ is a maximum (Fig. 56).

II. $f'(x)$ is negative for x less than x_0, and positive for x greater than x_0, i.e. $f'(x)$ passes through zero from negative to positive values.

In this case, we have an interval of decrease to the left of $x = x_0$, and one of increase to the right, i.e. $x = x_0$ corresponds to a peak of the curve that gives a minimum of the function (Fig. 56).

III. $f'(x)$ has the same sign for x either less than or greater than x_0. Suppose, for instance, that the sign is $(+)$.

The corresponding point of the graph in this case lies in an interval of increase, and is clearly not a peak (Fig. 57).

The above remarks lead us to the following rules for finding the values of x for which $f(x)$ has a maximum or minimum:

(1) find $f'(x)$;
(2) find the x for which $f'(x)$ is zero, i.e. solve the equation $f'(x) = 0$;

(3) find how the sign of $f'(x)$ varies on passing through these values, using the following arrangement:

x	x_0-h	x_0	x_0+h	$f(x)$
$f'(x)$	$+$	0	$-$	maximum
	$-$		$+$	minimum
	$+$		$+$	increasing
	$-$		$-$	decreasing

In the above table $x_0 - h$ and $x_0 + h$ indicate that the sign of $f'(x)$ is to be determined for x less than, and greater than, x_0, and always sufficiently close to x_0 so that h is taken as a sufficiently small positive number.

It is assumed in this investigation that $f'(x_0) = 0$, whilst $f'(x)$ differs from zero for all x sufficiently near, but not coinciding with, x_0.

We return to the fact that the tangent at M in Fig. 57 (with abscissa x_0) lies on different sides of the curve in the neighbourhood of M. Here, $f'(x_0) = 0$ and $f'(x) > 0$ for all x near to, but not at, x_0, all sections of the curve with x_0 as an interior point giving intervals of increase, notwithstanding the fact that $f'(x_0) = 0$.

Sometimes a rather different definition from the above is given of a maximum, viz., the function $f(x)$ has a maximum at $x = x_1$ if $f(x_1)$ is not less than the values of $f(x)$ at neighbouring points, i.e. if we have for the increment of the function, $f(x_1 + h) - f(x_1) \leqslant 0$, for all h sufficiently small in absolute value, and either positive or negative. A minimum at the point $x = x_2$ can be similarly defined by the inequality $f(x_2 + h) - f(x_2) \geqslant 0$. If the function in this definition has a derivative at its maximum or minimum, this derivative must be zero, as above.

To take an example, let it be required to find the maxima and minima of the function

$$f(x) = (x - 1)^2 (x - 2)^3.$$

We obtain the first derivative:

$$f'(x) = 2(x - 1)(x - 2)^3 + 3(x - 1)^2 (x - 2)^2 =$$

$$= (x - 1)(x - 2)^2 (5x - 7) = 5(x - 1)(x - 2)^2 \left(x - \frac{7}{5}\right).$$

It is clear from the last expression that $f'(x)$ is zero for the following values of the independent variable: $x_1 = 1$, $x_2 = 7/5$, $x_3 = 2$.

We now investigate these. For $x = 1$, the factor $(x - 2)^2$ has a plus sign, and $(x - 7/5)$ a minus sign. For all x sufficiently close to unity, and either greater than or less than unity, each of these factors keeps the same sign, so that their product is unconditionally minus for all x sufficiently near unity. We finally consider the factor $(x - 1)$, which is in fact zero at $x = 1$. It is negative for $x < 1$, and positive for $x > 1$. Thus the complete derivative, i.e. $f'(x)$, has a plus sign for $x < 1$ and a minus sign for $x > 1$. Hence it follows that $x = 1$ represents a maximum of the function $f(x)$. We set $x = 1$ in the expression for $f(x)$ itself, thus obtaining the value of the maximum, i.e. the ordinate of the corresponding peak of the graph of $f(x)$:

$$f(1) = 0^2 \cdot (-1)^3 = 0.$$

By using the same sort of argument for the other values $x_2 = 7/5$ and $x_3 = 2$, we obtain the following table:

x	$1-h$	1	$1+h$	$\frac{7}{5}-h$	$\frac{7}{5}$	$\frac{7}{5}+h$	$2-h$	2	$2+h$
$f'(x)$	$+$	0	$-$	$-$	0	$+$	$+$	0	$+$
$f(x)$	incr.	0 max.	decr.		$\frac{-108}{3125}$ min.	increases			

The method we have outlined for studying the maxima and minima of a function may have the drawback, especially in more complicated examples, that it is not too easy to find the sign of $f'(x)$ for x both greater than, and less than, the value in question. The trouble may be avoided in many cases by taking the second derivative $f''(x)$ into account. Suppose we consider $x = x_0$, for which $f'(x_0) = 0$. We shall assume that, on substituting $x = x_0$ in the expression for the second derivative, we obtain a positive value, i.e. $f''(x_0) > 0$. If $f'(x)$ is taken as the basic function, the fact that its derivative $f''(x)$ is positive at $x = x_0$ means that the basic function itself, $f'(x)$, is increasing at this point, i.e. $f'(x)$ passes from negative to positive values at its zero-point x_0. Thus, *for* $f''(x_0) > 0$ *at the point* $x = x_0$, $f(x)$ *will have a minimum. It can similarly be shown that, for* $f''(x_0) < 0$ *at* $x = x_0$, $f(x)$ *has a maximum.* Finally, if we get zero on substituting $x = x_0$ in the expression for $f''(x)$, i.e. $f''(x_0) = 0$, this prevents us from using the second derivative to investigate $x = x_0$, and we

must turn back to direct consideration of the sign of $f'(x)$. The arrangement shown in the table is thus obtained:

x	$f'(x)$	$f''(x)$	$f(x)$
x_0	0	$-$	maximum
		$+$	minimum
		0	doubtful case

It follows directly from the above discussion that, given the existence of the second derivative, a necessary condition for a maximum is $f''(x) \leqslant 0$, and a necessary condition for a minimum is $f''(x) \geqslant 0$. We can determine the maximum here by the condition $f(x_1 + h) - f(x_1) \leqslant 0$, and the minimum by $f(x_2 + h) - f(x_2) \geqslant 0$, as described above.

Example. It is required to find the maxima and minima of the function

$$f(x) = \sin x + \cos x.$$

This function has period 2π, i.e. remains unchanged on substituting $x + 2\pi$ for x.

It is sufficient to consider x varying in the interval $(0, 2\pi)$.

We find the first and second derivatives:

$$f'(x) = \cos x - \sin x; \; f''(x) = -\sin x - \cos x.$$

We put the first derivative equal to zero, obtaining the equation:

$$\cos x - \sin x = 0 \quad \text{or} \quad \tan x = 1.$$

The roots of this equation in $(0, 2\pi)$ are:

$$x_1 = \frac{\pi}{4} \text{ and } x_2 = \frac{5\pi}{4}.$$

We investigate these values by finding the sign of $f''(x)$:

$$f''\left(\frac{\pi}{4}\right) = -\sin\frac{\pi}{4} - \cos\frac{\pi}{4} = -\sqrt{2} < 0; \quad \text{maximum } f\left(\frac{\pi}{4}\right) = \sqrt{2};$$

$$f''\left(\frac{5\pi}{4}\right) = -\sin\frac{5\pi}{4} - \cos\frac{5\pi}{4} = \sqrt{2} > 0; \quad \text{minimum } f\left(\frac{5\pi}{4}\right) = -\sqrt{2}.$$

We note in conclusion a point that sometimes arises in finding maxima and minima. Points may occur on the graph of the function where there is either no tangent at all, or the tangent is parallel to OY (Fig. 58). The derivative $f'(x)$ will not exist at points of the first kind, whilst it is

infinite at points of the second kind, since the slope of a line parallel to OY is infinity. As is evident at once from the figure, however, these points can represent maxima or minima of the function. The above rule for finding maxima and minima should thus be amplified as follows, strictly speaking: *maxima and minima of a function $f(x)$ can occur not only at points where $f'(x)$ is zero, but also at points where $f'(x)$ either does not exist or becomes infinite.* Investigation of these latter points must be carried out by the first of the arrangements given above, i.e. by finding the sign of $f'(x)$ for x less than, and greater than, the value in question.

Example. It is required to find the maxima and minima of

$$f(x) = (x-1)\sqrt[3]{x^2}.$$

We find the first derivative:

$$f'(x) = \sqrt[3]{x^2} + \frac{2(x-1)}{3\sqrt[3]{x}} = \frac{5}{3} \times \frac{x - \frac{2}{5}}{\sqrt[3]{x}}.$$

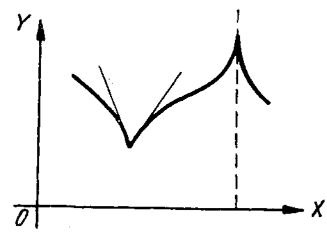

FIG. 58

This is zero for $x = 2/5$ and infinity for $x = 0$. We consider the latter value: the numerator of the above fraction is minus for zero x, and also minus for all positive or negative x that are near zero. The denominator of the fraction is negative for $x < 0$, and positive for $x > 0$. The fraction as a whole is therefore positive for x less than, and near, zero, and negative for $x > 0$, i.e. we have a maximum at $x = 0$, $f(0) = 0$. We have a minimum at $x = 2/5$:

$$f\left(\frac{2}{5}\right) = -\frac{3}{5}\sqrt[3]{\frac{4}{25}} = -\frac{3}{25}\sqrt[3]{20}.$$

59. Curve tracing. Finding the maxima and minima of a function $f(x)$ considerably simplifies the problem of tracing its curve. This is explained by means of a few examples.

1. Let it be required to trace the graph of:

$$y = (x-1)^2(x-2)^3,$$

which we considered in a previous paragraph. We then obtained two peaks of the function: a maximum $(1,0)$ and a minimum $(7/5, -108/3125)$. We fill in these points in the figure. It is also useful to mark the intercepts of the curve on the axes. We have for $x = 0$, $y = -8$, i.e. the point on axis OY is $y = -8$.

We put y equal to zero, i.e.,

$$(x-1)^2(x-2)^3 = 0,$$

and obtain the points on axis OX. One of these, $x=1$, is a peak, as already explained, whilst the second, $x = 2$, is not a peak, as was also explained earlier, but corresponds to a point of the graph where the tangent is parallel to OX. The required curve is shown in Fig. 59.

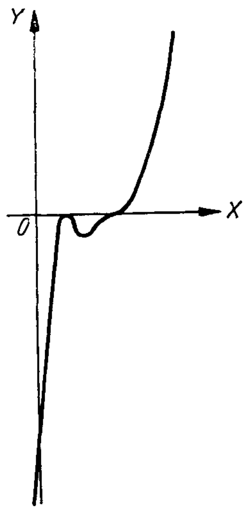

FIG. 59

2. We trace the curve:

$$y = e^{-x^2}.$$

We find the first derivative:

$$y' = -2xe^{-x^2}.$$

Putting $y' = 0$, we get the value $x = 0$, corresponding to a peak (maximum) of the curve, as is easily seen, with ordinate $y = 1$. This also gives us the point of the curve on OY. Putting y zero, we get the equation $e^{-x^2} = 0$, which has no solution, i.e. the curve has no points on OX. But we note that as x tends to $(+\infty)$ or $(-\infty)$, the power in e^{-x^2} tends to $-\infty$ and the expression as a whole tends to zero, i. e. on infinite displacement to left or right, the curve indefinitely approaches the axis OX. The curve is shown in Fig. 60, in accordance with the data obtained.

3. We draw the curve

$$y = e^{-ax} \sin bx \quad (a > 0),$$

representing a *damped oscillation*. The factor $\sin bx$ does not exceed unity in absolute value, and the graph as a whole is situated between the two curves:

$$y = e^{-ax} \quad \text{and} \quad y = -e^{-ax}.$$

As x tends to $(+\infty)$, factor e^{-ax}, and hence the product $e^{-ax} \sin bx$ as a whole, tends to zero, i.e. the curve approaches OX asymptotically on infinite displacement to the right. The points of the curve on OX are given by the equation

$$\sin bx = 0,$$

so that they are

$$x = \frac{k\pi}{b} \quad (k \text{ is an integer}).$$

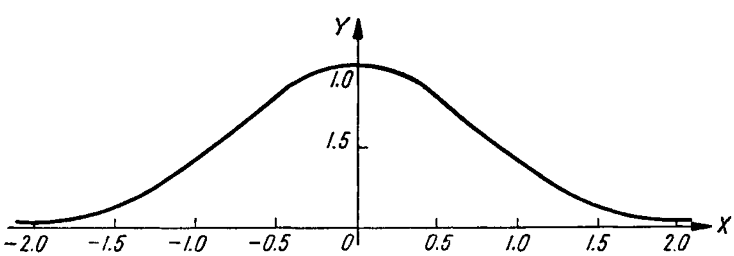

FIG. 60

We find the first derivative:

$$y' = -ae^{-ax}\sin bx + be^{-ax}\cos bx = e^{-ax}(b\cos bx - a\sin bx).$$

The expression in brackets can be put in the well-known form:

$$b\cos bx - a\sin bx = K\sin(bx + \varphi_0),$$

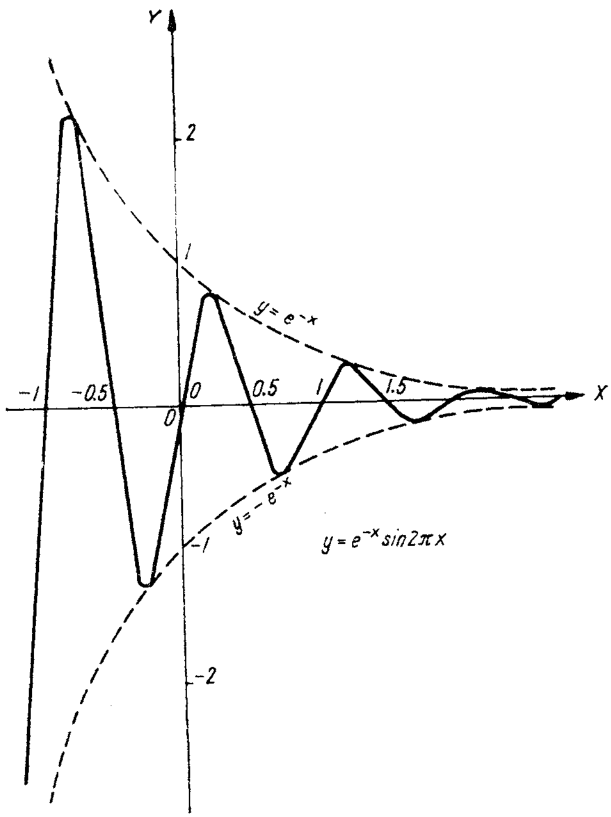

FIG. 61

where K and φ_0 are constants. On putting the first derivative equal to zero we get the equation:

$$\sin(bx + \varphi_0) = 0,$$

which gives:

$$bx + \varphi_0 = k\pi, \quad \text{i.e.} \quad x = \frac{k\pi - \varphi_0}{b} \quad (k \text{ is an integer}) \qquad (1)$$

When x passes through these values, $\sin(bx + \varphi_0)$ changes its sign each time. The same can evidently be said of the derivative y', since,

$$y' = Ke^{-ax}\sin(bx + \varphi_0),$$

and the sign of the factor e^{-ax} is constant. There are thus alternate maxima and minima corresponding to these roots. We should have a sine wave without the exponential factor:

$$y = \sin bx ,$$

and the abscissae of its peaks would be given by:

$$\cos bx = 0 ,$$

i.e.

$$x = \frac{(2k-1)\pi}{2b} \quad (k \text{ an integer}). \tag{1_1}$$

We thus see that the exponential factor not only reduces the amplitude of the oscillation, but also displaces the abscissae of the peaks of the curve. It is evident, on comparing equations (1) and (1_1), that this displacement is equal to the constant $(-\pi/2b - \varphi_0/b)$. The graph of the damped oscillation is shown in Fig. 61 for $a = 1$ and $b = 2\pi$. The peaks of the curve do not lie on the dotted lines corresponding to $y = \pm e^{-ax}$. This is due to the above mentioned displacement of the peaks.

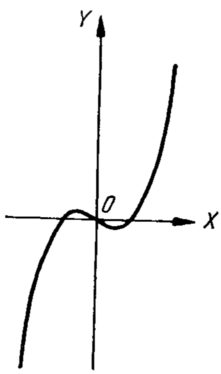

FIG. 62

4. We draw the curve:

$$y = \frac{x^3 - 3x}{6} .$$

We find the first and second derivatives:

$$y' = \frac{x^2 - 1}{2} ; \quad y'' = x.$$

We put the first derivative equal to zero, and find the two values $x_1 = 1$ and $x_2 = -1$. Substituting these values in the second derivative, we see that the first value represents a minimum, and the second a maximum. We substitute these values in the expression for y, finding the corresponding peaks of the curve:

$$\left(-1, \ \frac{1}{3}\right), \ \left(1, \ -\frac{1}{3}\right).$$

Putting $x = 0$, we get $y = 0$, i.e. the origin (0,0) lies on the curve. Finally, we get for zero y, apart from $x = 0$, two further points $x = \pm\sqrt{3}$, so that the curve cuts the axes in three points, (0,0), $(\sqrt{3},0)$ and $(-\sqrt{3},0)$. We note further that, on simultaneously replacing x and y by $-x$ and $-y$, both sides of the equation of the curve only change sign, i.e. the origin is a centre of symmetry of the curve (Fig. 62).

60. The greatest and least values of a function. We assume that the values of the function $f(x)$ are considered for values of the independent

variable x in the interval (a, b); it is required to find the greatest and the least of these values. If $f(x)$ is continuous, it will certainly attain a greatest and least value, as we saw in [35], so that the corresponding graph will have a greatest and a least ordinate in the interval in question. We can find all the maxima and minima of the function inside the interval (a, b) in accordance with the rule given above. If the function $\varphi(x)$ has its a greatest ordinate inside this interval, this will evidently coincide with the greatest maximum of the function inside the interval. It can happen, however, that the greatest ordinate is not inside the interval, but at one of the ends $x = a$ or $x = b$. Hence it is not sufficient, in finding say the greatest value of a function, to compare all its maxima inside the interval and take the greatest, since its values at the ends of the interval must also be taken into account. Similarly, the least value of a function must be found by taking all its minima inside the interval, together with the boundary values of the function or $x = a$ and $x = b$. We remark here that maxima and minima can be completely absent, yet a greatest and least value must exist for a continuous function in a bounded interval (a, b).

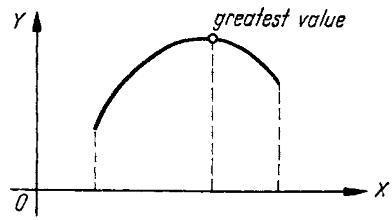

FIG. 63

We note some particular cases, when the greatest and least values can be found very simply. If, for example, $f(x)$ is increasing in (a, b), it will obviously take its least value at $x = a$, and its greatest value at $x = b$. The opposite will be the case for $f(x)$ decreasing.

If the function has a single maximum inside the interval, with no minima, this single maximum gives the greatest value of the function (Fig. 63), so that it is quite unnecessary to find the values at the ends of the interval in order to find the greatest value of the function in this case. Similarly, if the function has a single minimum inside the interval, with no maxima, the minimum gives the least value of the function. The cases just mentioned are met with in the first of the four examples given below.

1. *It is required to cut a given line, of length l, into two parts, such that the rectangle formed from these parts has a maximum area.*

Let x be the length of one part of the line, so that the length of the other part is $(l - x)$. Since the area of a rectangle is the product of its neighbouring

sides, we see that the problem reduces to finding the value of x in the interval $(0, l)$ for which the function:

$$f(x) = x(l - x)$$

attains its greatest value.

We find the first and second derivatives:

$$f'(x) = (l - x) - x = l - 2x; \; f''(x) = -2 < 0.$$

Equating the first derivative to zero, we get a single value $x = l/2$, which corresponds to a maximum, since $f''(x)$ is a negative constant. The greatest area is thus obtained with a square of side $l/2$.

2. *A sector is cut out from a circle of radius R, and the remainder is glued together into a cone. It is required to find for what angle of the sector cut out the volume of the cone has its greatest value.*

Instead of taking the angle of the sector cut out as the independent variable, we take the difference between this angle and 2π as x, i.e. x is the angle of the remaining sector. For x near 0 or 2π, the volume of the cone approaches zero, so that there is evidently a value of x in the interval $(0, 2\pi)$ for which the volume is greatest.

The cone obtained by glueing together the remainder of the circle must have a generator of length R, a length of circular base Rx, a radius of base $r = Rx/2\pi$, and height:

$$h = \sqrt{R^2 - \frac{R^2 x^2}{4\pi^2}} = \frac{R}{2\pi} \sqrt{4\pi^2 - x^2}.$$

The volume of this cone will be:

$$v(x) = \frac{1}{3} \pi \frac{R^2 x^2}{4\pi^2} \cdot \frac{R}{2\pi} \sqrt{4\pi^2 - x^2} = \frac{R^3}{24\pi^2} x^2 \cdot \sqrt{4\pi^2 - x^2}.$$

The constant factor $R^3/24\pi^2$ can be neglected when finding the greatest value of this function. The remaining product $x^2 \sqrt{4\pi^2 - x^2}$ is positive, and therefore attains its greatest value for the same x as that for which its square attains its greatest value. We can thus consider the function:

$$f(x) = 4\pi^2 x^4 - x^6$$

inside the interval $(0,2\pi)$.

We find the first derivative:

$$f'(x) = 16\pi^2 x^3 - 6x^5.$$

It exists for all x. Equating it to zero, we obtain three values:

$$x_1 = 0, \quad x_2 = -2\pi \sqrt{\frac{2}{3}}, \; x_3 = 2\pi \sqrt{\frac{2}{3}}.$$

The first two values *do not lie inside* the interval $(0,2\pi)$. The single value $x_3 = 2\pi \sqrt{2/3}$ remains, inside the interval; and since we saw above that the x giving the greatest volume of cone must lie inside the interval, we can say without investigating x_3 that this is the required x.

3. *A straight line L divides a plane into two parts (media) I and II. A point moves in medium I with speed v_1, and in medium II with speed v_2. In what path must a point move in order to pass as quickly as possible from point A in medium I to point B in medium II?*

Let AA_1 and BB_1 be the perpendiculars from A and B onto L. We introduce the following notation:

$$\overline{AA_1} = a, \quad \overline{BB_1} = b, \quad \overline{A_1B_1} = c,$$

and the abscissae will be measured on L in the direction $\overline{A_1B_1}$ (Fig. 65).

It is obvious that the point must have a straight-line path in both media, I and II, though generally speaking, line AB will not represent the "quickest path". The "quickest path" will thus consist of two straight sections $\overline{AM}$

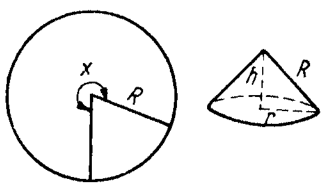

FIG. 64

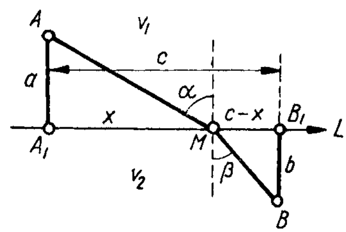

FIG. 65

and $\overline{MB}$, with point M lying on L. We take the abscissa of M as independent variable x ($x = \overline{A_1 M}$). We want the least value for time t, given by:

$$t = f(x) = \frac{AM}{v_1} + \frac{MB}{v_2} = \frac{\sqrt{a^2 + x^2}}{v_1} + \frac{\sqrt{b^2 + (c - x)^2}}{v_2}$$

in the interval $(-\infty, +\infty)$.

We find the first and second derivatives:

$$f'(x) = \frac{x}{v_1\sqrt{a^2 + x^2}} - \frac{c - x}{v_2\sqrt{b^2 + (c - x)^2}}$$

$$f''(x) = \frac{a^2}{v_1(a^2 + x^2)^{3/2}} + \frac{b^2}{v_2[b^2 + (c - x)^2]^{3/2}}$$

Both derivatives exist for all x, and $f''(x)$ is always positive. Hence, $f'(x)$ is increasing in the interval $(-\infty, +\infty)$, and cannot become zero more than once. But

$$f'(0) = -\frac{c}{v_2\sqrt{b^2 + c^2}} < 0$$

and

$$f'(c) = \frac{c}{v_1\sqrt{a^2 + c^2}} > 0,$$

and hence the equation

$$f'(x) = 0$$

has a single root x_0 between 0 and c, giving the single minimum of $f(x)$, since $f''(x) > 0$. The abscissae 0 and c represent points A_1 and B_1, so that the required point M will lie between A_1 and B_1, as could also be shown by elementary geometry.

We give the geometrical interpretation of the solution obtained. Let α and β denote the angles formed by segment $\overline{AM}$ and $\overline{BM}$ respectively with the perpendicular to L at M. The abscissa x of the required point M must make $f'(x)$ zero, i.e. must satisfy the equation:

$$\frac{x}{v_1\sqrt{a^2+x^2}} = \frac{c-x}{v_2\sqrt{b^2+(c-x)^2}},$$

which can be rewritten as:

$$\frac{A_1 M}{v_1 \cdot AM} = \frac{MB_1}{v_2 \cdot BM}$$

or

$$\frac{\sin\alpha}{v_1} = \frac{\sin\beta}{v_2}, \quad \text{i.e.} \quad \frac{\sin\alpha}{\sin\beta} = \frac{v_1}{v_2};$$

"the quickest path" will be that for which the ratio of the sines of the angles α and β is equal to the ratio of the speeds in media I and II. This result gives us the well-known law for the refraction of light, so that refraction takes place with the ray of light choosing the "quickest path" from a point in one medium to a point in another.

4. We suppose that the experimental determination of a magnitude x involves making n individual careful observations, giving n values:

$$a_1, a_2, \ldots, a_n,$$

which are not identical due to instrument inaccuracies. The value of x giving the least sum of squares of errors is taken as the "best" value. Finding this best value thus means finding the x giving the least value of the function

$$f(x) = (x-a_1)^2 + (x-a_2)^2 + \ldots + (x-a_n)^2$$

in the interval $(-\infty, +\infty)$.

We find the first and second derivatives:

$$f'(x) = 2(x-a_1) + (x-a_2) + \ldots + 2(x-a_n),$$

$$f''(x) = 2 + 2 + \ldots + 2 = 2n > 0.$$

Making the first derivative zero, we get a single value:

$$x = \frac{a_1 + a_2 + \ldots + a_n}{n},$$

which represents a minimum, since the second derivative is positive. Thus, *the "best" value of x is the arithmetic mean of the observational results.*

5. *To find the shortest distance from a point M to a circle.*

We take the centre of the circle as origin O, and the line OM as axis OX. Let $OM = a$, and let R be the radius of the circle. The equation of the circle is:

$$x^2 + y^2 = R^2 ,$$

and the distance of M, with coordinates $(a, 0)$, to any point of the circle will be:

$$\sqrt{(x - a)^2 + y^2}.$$

We shall find the least value of the square of this distance. On substituting for y^2 from the equation of the circle, we get the function :

$$f(x) = (x - a)^2 + (R^2 - x^2) = -2ax + a^2 + R^2 ,$$

where the independent variable x lies in the interval $(-R \leqslant x \leqslant R)$.

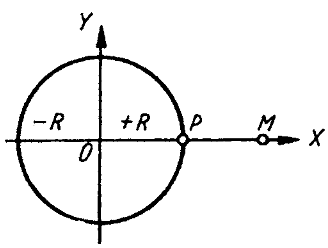

FIG. 66

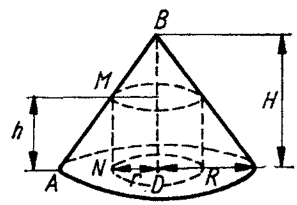

FIG. 67

Since the first derivative $f'(x) = -2a$ is negative for all x, function $f(x)$ is decreasing, and hence attains its least value for $x = R$ at the right-hand end of the interval. The shortest distance is thus given by $\overline{PM}$ (Fig. 66).

6. *Given a right circular cone, it is required to inscribe in it a cylinder with the greatest total surface area.*

Let the radius of base of the cone be R and its height H, and let the radius of base and height of the cylinder be r and h. The function whose greatest value is required is here:

$$S = 2\pi r^2 + 2\pi rh .$$

Variables r and h are connected, due to the fact of inscribing the cylinder in the given cone. We have from the similar triangles ABD and AMN (Fig. 67):

$$\frac{MN}{AN} = \frac{BD}{AD} ,$$

or

$$\frac{h}{R - r} = \frac{H}{R} ,$$

whence

$$h = \frac{R - r}{R} H .$$

We obtain on substituting for h in the expression for S:

$$S = 2\pi\left[r^2 + rH\left(1 - \frac{r}{R}\right)\right].$$

Hence, S is a function of the single independent variable r, which must lie in the interval $0 \leqslant r \leqslant R$. We find the first and second derivatives:

$$\frac{dS}{dr} = 2\pi\left(2r + H - \frac{2r}{R}H\right), \quad \frac{d^2S}{dr^2} = 4\pi\left(1 - \frac{H}{R}\right).$$

We find a single value of r, on making dS/dr zero:

$$r = \frac{HR}{2(H-R)}. \tag{2}$$

For this value to lie inside $(0, R)$, the inequalities must be satisfied:

$$0 < \frac{HR}{2(H-R)} \quad \text{and} \quad \frac{HR}{2(H-R)} < R. \tag{3}$$

The first inequality is equivalent to the condition that $H > R$. We multiply both sides of the second inequality by the positive term $2(H - R)$ and get:

$$R < \frac{1}{2}H.$$

With fulfilment of this condition, d^2S/dr^2 is negative; the value (2) gives the single maximum of function S, and the greatest surface-area of the cylinder. The size of the latter can easily be found by substituting for r from (2) in the expression for S.

We now take the case of value (2) not lying inside $(0, R)$ i.e. one of the inequalities (3) is not satisfied. Two possibilities can now arise: either $H < R$ or $H > R$, but $R \geqslant H/2$. Both these can be characterized by the single inequality:

$$H \leqslant 2R. \tag{4}$$

We rewrite dS/dr as:

$$\frac{dS}{dr} = 2\pi\left(2r + H - \frac{2r}{R}H\right) = \frac{2\pi}{R}[(2R - H)r + H(R - r)].$$

It is clear from the new expression that $dS/dr > 0$ for $0 < r < R$ on satisfying (4), i.e. S is increasing in $(0, R)$, and thus attains its greatest value for $r = R$. Evidently, $h = 0$ with this value of r, and we can look on the result obtained as a flattened cylinder, the base of which coincides with the base of the cone, and the total surface-area of which gives $2\pi R^2$.

61. Fermat's theorem. We have used elementary geometry above, to give methods for studying the increase and decrease of a function, as well as for finding its maxima and minima and its greatest and least values. We now turn to the rigorous analytic statement of some theorems and formulae, which give us an analytic proof of the rules found above, and also enable the study of functions

to be carried somewhat further. When stating the next theorems and formulae, we shall include a detailed and precise account of all the conditions for which they are valid.

FERMAT'S THEOREM. *If $f(x)$ is continuous in (a, b), has a derivative at every interior point of the interval, and attains its greatest (or least) value at some interior point $x = c$, the first derivative must be zero at* $x = c$, i.e. $f'(c) = 0$.

We suppose for clarity that $f(c)$ is the greatest value of the function. The proof will be exactly similar, in the case of its being the least value. In accordance with the condition that $x = c$ lies inside the interval, the difference $f(c + h) - f(c)$ will be negative, or at any rate not positive, for any positive or negative h:

$$f(c + h) - f(c) \leqslant 0.$$

We take the ratio:

$$\frac{f(c+h) - f(c)}{h}.$$

The numerator of this fraction is less than or equal to zero, as remarked, so that:

$$\frac{f(c+h) - f(c)}{h} \leqslant 0 \text{ for } h > 0, \tag{5}$$

$$\frac{f(c+h) - f(c)}{h} \geqslant 0 \text{ for } h < 0.$$

Point $x = c$ lies inside the interval, and a derivative exists at the point by hypothesis, i.e. the above fraction tends to a definite limit $f'(c)$, when h tends to zero in any manner. We first suppose that h tends to zero through positive values. Passing to the limit in the first of inequalities (5), we get:

$$f'(c) \leqslant 0.$$

Similarly, passing to the limit with $h \to 0$ in the second inequality of (5) gives:

$$f'(c) \geqslant 0.$$

Taking these two inequalities together, we obtain the required result:

$$f'(c) = 0.$$

62. Rolle's theorem. *If $f(x)$ is continuous in (a, b), has a derivative at every interior point of the interval, and has equal values at the ends of the interval,* i.e. $f(a) = f(b)$, *then there exists at least one interior point, $x = c$, such that the derivative is zero,* i.e. $f'(c) = 0$.

A continuous function $f(x)$ must havee a last value m and a greatest value M in the interval considered. If the least and greatest values happened to coincide, i.e. $m = M$, it would follow that the function kept a constant value m (or M) throughout the interval. We know that the derivative of a constant is zero, so that in this simple case, the derivative would be zero at every interior point of the interval. Turning to the general case, we can thus suppose that $m < M$. Since the function has the same value at the limits, i.e. $f(a) = f(b)$, by hypothesis, at least one of the numbers m or M must differ from this common value. Suppose, for example, that it is M, i.e. the greatest value of the function is reached inside, and not at the limits, of the interval. Let $x = c$ be the point at which this value is reached. By Fermat's theorem, we shall have $f'(c) = 0$ at this point, which proves Rolle's theorem.

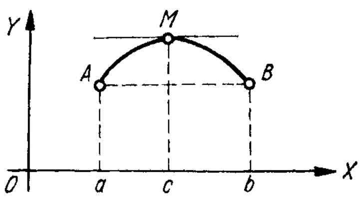

FIG. 68

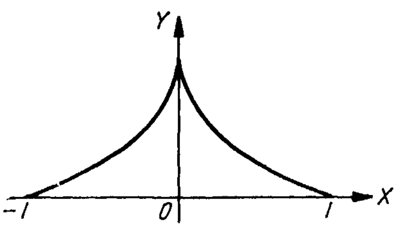

FIG. 69

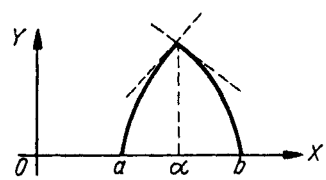

FIG. 70

In the particular case of $f(a) = f(b) = 0$, Rolle's theorem can be briefly formulated as: *the first derivative of $f(x)$ will vanish at at least one point in (a, b).*

There is a simple geometrical interpretation of Rolle's theorem. By hypothesis, $f(a) = f(b)$, i.e. the ordinates of the curve $y = f(x)$ are equal at the ends of the interval, and a derivative exists inside the interval, i.e. the curve has a definite tangent. Rolle's theorem says that there exists at least one interior point in this case, where the derivative is zero, i.e. where the tangent is parallel to axis OX (Fig. 68).

Remark. If the condition regarding the existence of the derivative $f'(x)$ is not fulfilled for every interior point of the interval in Rolle's theorem, the theorem can be proved false.

Take, for example,

$$f(x) = 1 - \sqrt[3]{x^2};$$

this is continuous in the interval $(-1, 1)$, and $f(-1) = f(1) = 0$, but the derivative,

$$f'(x) = -\frac{2}{3\sqrt[3]{x}}$$

does not vanish inside the interval. This follows from the fact that $f'(x)$ does not exist (tends to infinity) for $x = 0$ (Fig. 69) Another example is the curve shown in Fig. 70. Here we have the curve $y = f(x)$, with $f(a) = f(b) = 0$. But it is evident from the figure that the tangent cannot be parallel to OX inside the interval (a, b), i.e. $f'(x)$ does not vanish. This happens because the curve has two different tangents at $x = a$, to the left and right of this point, and hence a unique derivative does not exist at this point, so that the condition of Rolle's theorem regarding the existence of the derivative at every interior point is not fulfilled.

63. Lagrange's formula. We assume that $f(x)$ is continuous in (a, b), and has a derivative inside this interval, but that the condition $f(a) = f(b)$ of Rolle's theorem cannot be satisfied. We form a new function:

$$F(x) = f(x) + \lambda x,$$

where λ is a constant, so defined that the new function satisfies the condition mentioned for Rolle's theorem, i.e. we make:

$$F(a) = F(b),$$

so that

$$f(a) + \lambda a = f(b) + \lambda b,$$

whence

$$\lambda = -\frac{f(b) - f(a)}{b - a}.$$

Applying Rolle's theorem now to $F(x)$, we can say that there will be a point $x = c$ between a and b, where

$$F'(c) = f'(c) + \lambda = 0 \quad (a < c < b),$$

whence, on substituting the above expression for λ:

$$f'(c) = -\lambda \quad \text{or} \quad f'(c) = \frac{f(b) - f(a)}{b - a}.$$

The last equation can be written:

$$f(b) - f(a) = (b - a)\, f'(c).$$

This equation is called *Lagrange's formula*. The value c lies between a and b, so that the ratio $(c-a)/(b-a) = \theta$ lies between zero and unity, and we can write:

$$c = a + \theta(b - a) \quad (0 < \theta < 1),$$

and Lagrange's formula now takes the form:

$$f(b) - f(a) = (b - a) f'(a + \theta(b - a)) \quad (0 < \theta < 1).$$

Putting $b = a + h$, the formula becomes:

$$f(a + h) - f(a) = hf'(a + \theta\, h).$$

Lagrange's formula gives an accurate expression for the increment $f(b) - f(a)$ of the function $f(x)$, so that it is also called the *formula of finite increments.*

We know that the derivative of a constant is zero. Lagrange's formula allows us to state the converse: *if the derivative $f'(x)$ is zero for every point of (a, b), $f(x)$ is constant in the interval.*

In fact, taking an arbitrary x of (a, b), and applying Lagrange's formula to the interval (a, x), we get:

$$f(x) - f(a) = (x - a)f'(\xi) \quad (a < \xi < x);$$

but $f'(\xi) = 0$ by hypothesis, hence:

$$f(x) - f(a) = 0, \quad \text{i.e.} \quad f(x) = f(a) = \text{constant}.$$

All we know as regards the magnitude c that appears in Lagrange's formula is that it lies between a and b, so that the formula does not enable us accurately to calculate the increment of a function from its derivative; the formula can be used, however, to estimate the error involved in replacing the increment of a function by its differential.

Example. Let

$$f(x) = \log_{10} x.$$

The derivative is

$$f'(x) = \frac{1}{x} \cdot \frac{1}{\log 10} = \frac{M}{x}\,(M = 0 \cdot 43429 \ldots),$$

and Lagrange's formula gives us:

$$\log_{10}(a+h) - \log_{10} a = h \frac{M}{a+\theta h} \quad (0 < \theta < 1)$$

or

$$\log_{10}(a+h) = \log_{10} a + h \frac{M}{a+\theta h}.$$

Replacing the increment by the differential gives us the approximate formula:

$$\log_{10}(a+h) - \log_{10} a = h \frac{M}{a}, \quad \log_{10}(a+h) = \log_{10} a + h \frac{M}{a}.$$

On comparing this approximate equation with the accurate one, obtained with Lagrange's formula, we see that the error is:

$$h\frac{M}{a} - h\frac{M}{a+\theta h} = \frac{\theta h^2 M}{a(a+\theta h)}.$$

Putting $a = 100$ and $h = 1$, we get the approximate equation:

$$\log_{10} 101 = \log_{10} 100 + \frac{M}{100} = 2.00434\ldots$$

with the error

$$\frac{\theta \times M}{100\,(100+\theta)} \qquad (0 < \theta < 1).$$

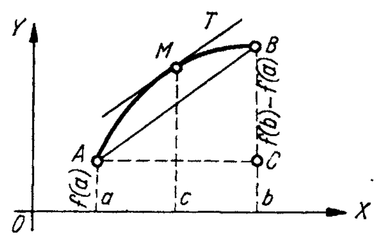

FIG. 71

Replacing θ by unity in the numerator, and by zero in the denominator, of this fraction, we can say on evaluating the fraction that the error in calculating the value of $\log_{10} 101$ is less than

$$\frac{M}{100^2} = 0.00004\ldots$$

We can rewrite Lagrange's formula as:

$$\frac{f(b) - f(a)}{b-a} = f'(c) \qquad (a < c < b).$$

We see from the graph of $y = f(x)$ (Fig. 71) that the ratio

$$\frac{f(b) - f(a)}{b-a} = \frac{\overline{CB}}{\overline{AC}} = \tan \angle CAB$$

gives slope of the chord AB, while $f'(c)$ gives the slope of the tangent at some point M of the segment AB of the curve. Lagrange's formula thus amounts to the assertion: there is a point in a segment of a curve where the tangent is parallel to the chord. Rolle's theorem

is a special case of this assertion, when the chord is parallel to OX, i.e. $f(a) = f(b)$.

Remark. The tests for increase and decrease, which we established above from the graph, follow immediately from Lagrange's formula. We suppose that the first derivative is positive inside a certain interval, and we let x and $x + h$ be two points of this interval. Evidently, from Lagrange's formula:

$$f(x + h) - f(x) = hf'(x + \theta h) \quad (0 < \theta < 1)\,,$$

the difference on the left will be positive for positive h, since both factors of the product on the right are positive in this case. The assumption that the derivative is positive in some interval thus leads us to:

$$f(x + h) - f(x) > 0\,,$$

i.e. the function is increasing in this interval. The test for a decreasing function similarly follows directly from the formula written above.

We also note here that the argument used in proving Fermat's theorem is fully applicable to the case when the function reaches a maximum or minimum, and not necessarily its greatest or least value, at the point in question. This argument shows us that the first derivative must be zero at such points, if it exists.

64. Cauchy's formula. We take $f(x)$ and $\varphi(x)$ continuous in (a, b), with derivatives at every interior point of the interval, and with $\varphi'(x)$ not zero at any interior point. We apply Lagrange's formula to $\varphi(x)$, giving

$$\varphi(b) - \varphi(a) = (b - a)\varphi'(c_1) \qquad (a < c_1 < b),$$

with $\varphi(c_1) \neq 0$ by hypothesis, whence

$$\varphi(b) - \varphi(a) \neq 0.$$

We form the function:

$$F(x) = f(x) + \lambda\varphi(x)\,,$$

where λ is a constant, so defined that $F(a) = F(b)$, i.e.

$$f(a) + \lambda\varphi(a) = f(b) + \lambda\varphi(b)\,,$$

whence

$$\lambda = -\frac{f(b) - f(a)}{\varphi(b) - \varphi(a)}\,.$$

Rolle's theorem is applicable to $F(x)$ with this choice of λ, giving us the existence of $x = c$ such that

$$F'(c) = f'(c) + \lambda\varphi'(c) = 0 \qquad (a < c < b).$$

This equation gives:

$$\frac{f'(c)}{\varphi'(c)} = -\lambda \qquad (\varphi'(c) \neq 0),$$

whence, on substituting the value found for λ, we get:

$$\frac{f(b) - f(a)}{\varphi(b) - \varphi(a)} = \frac{f'(c)}{\varphi'(c)} \qquad (a < c < b)$$

or

$$\frac{f(b) - f(a)}{\varphi(b) - \varphi(a)} = \frac{f'[a + \theta(b - a)]}{\varphi'[a + \theta(b - a)]} \qquad (0 < \theta < 1), \qquad (6)$$

or

$$\frac{f(a + h) - f(a)}{\varphi(a + h) - \varphi(a)} = \frac{f'(a + \theta h)}{\varphi'(a + \theta h)}.$$

This is Cauchy's formula. On putting $\varphi(x) = x$ in this formula, we have $\varphi'(x) = 1$, and the formula becomes:

$$\frac{f(b) - f(a)}{b - a} = \frac{f'(c)}{1}$$

or

$$f(b) - f(a) = (b - a)f'(c),$$

i.e. we get Lagrange's formula as a special case of Cauchy's formula.

65. Evaluating indeterminate forms. If two functions $\varphi(x)$ and $\psi(x)$ vanish at $x = a$, the fraction $\varphi(x)/\psi(x)$ is an indeterminate form of the type 0/0 at $x = a$. We indicate a method of evaluating such indeterminate forms. We suppose that $\varphi(x)$ and $\psi(x)$ are continuous and have a first derivative near $x = a$, whilst $\psi'(x)$ does not vanish for x near, but not equal to, a.

We prove the following theorem: *if, with the above assumptions, $\varphi'(x)/\psi'(x)$ tends to a limit b as x tends to a, then $\varphi(x)/\psi(x)$ tends to the same limit.*

Noting that:

$$\psi(a) = \varphi(a) = 0,$$

and using Cauchy's formula [64], we get:

$$\frac{\varphi(x)}{\psi(x)} = \frac{\varphi(x) - \varphi(a)}{\psi(x) - \psi(a)} = \frac{\varphi'(\xi)}{\psi'(\xi)} \qquad (\xi \text{ between } a \text{ and } x). \qquad (7)$$

We remark that we are justified in applying Cauchy's formula, given the assumptions made regarding $\varphi(x)$ and $\psi(x)$.

If x tends to a, ξ, lying between x and a, will tend to the same limit. The right-hand side of (7) now tends to b, by hypothesis, and hence $\varphi(x)/\psi(x)$ on the left-hand side of (7) tends to the same limit.

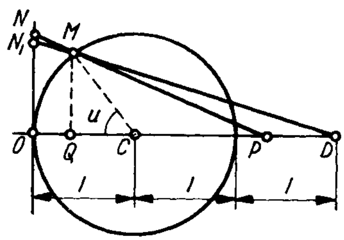

FIG. 72

The theorem just proved gives us the following rule for evaluating an indeterminacy of the form 0/0:

To find the limit of $\varphi(x)/\psi(x)$, in the case of an indeterminacy of the form 0/0, the ratio of the functions can be replaced by the ratio of the derivatives, and the limit found of this new ratio.

This rule was given by the French mathematician l'Hôpital, and is usually named after him.

If the ratio of the derivatives also leads to an indeterminacy of the form 0/0, the rule can be applied to this ratio, and so on.

We give some examples of using the rule:

1. $$\lim_{x\to 0} \frac{(1+x)^n - 1}{x} = \lim_{x\to 0} \frac{n(1+x)^{n-1}}{1} = n.$$

2. $$\lim_{x\to 0} \frac{x - \sin x}{x^3} = \lim_{x\to 0} \frac{1-\cos x}{3x^2} = \lim_{x\to 0} \frac{\sin x}{6x} = \lim_{x\to 0} \frac{\cos x}{6} = \frac{1}{6},$$

i.e. the difference $x - \sin x$ is an infinitesimal of the third order with respect to x.

3. $$\lim_{x\to 0} \frac{x - x\cos x}{x - \sin x} = \lim_{x\to 0} \frac{1-\cos x + x\sin x}{1-\cos x} =$$
$$= \lim_{x\to 0} \frac{\sin x + \sin x + x\cos x}{\sin x} = \lim_{x\to 0} \frac{2\cos x + \cos x - x\sin x}{\cos x} = 3.$$

The result of this example leads to a practical method for *rectifying the arc of a circle.*

We take a circle of radius unity. A diameter of the circle is taken as OX, and the tangent at one end of the diameter as OY. (See Fig. 72.)

We take a certain arc OM, and let ON be an interval along OY equal to arc OM. We draw NM and let P be its point of intersection with OX.

Let u denote the length of arc OM (radius taken as unity). The equation of NM can be written in terms of its intersections with the axes:

$$\frac{x}{OP} + \frac{y}{u} = 1.$$

We find the length of $\overline{OP}$ by noting that M lies on NM and has the coordinates:

$$x = OQ = 1 - \cos u,\ y = QM = \sin u.$$

These coordinates must satisfy the following equation:

$$\frac{1 - \cos u}{OP} + \frac{\sin u}{u} = 1,$$

whence

$$OP = \frac{u - u \cos u}{u - \sin u}.$$

The result of Example 3 shows that $OP \to 3$ as $u \to 0$, i.e. P tends to D on OX, the distance of D from the origin being three times the radius. This gives a simple method of approximate rectification of the arc of a circle. To rectify arc OM, section $\overline{OD}$ must be taken from O, three times the radius of the circle, then the line DM must be drawn. The intercept of DM on OY, $\overline{ON_1}$, gives the approximate length of arc OM. This method gives a very satisfactory result, especially for small arcs; even for an arc of $\pi/2$, the relative error is about 5%.

66. Other indeterminate forms. The theorem of [65] can be justified for an indeterminacy of the form $^{\infty}/_{\infty}$. Let

$$\lim_{x\to a} \varphi(x) = \lim_{x\to a} \psi(x) = \infty \tag{8}$$

and

$$\lim_{x\to a} \frac{\varphi'(x)}{\psi'(x)} = b. \tag{9}$$

We show that $\varphi(x)/\psi(x)$ tends to the same limit b, assuming that $\psi'(x)$ does not vanish for x near a.

We consider two values x and x_0 of the independent variable, near a, and such that x lies between x_0 and a. We have by Cauchy's formula:

$$\frac{\varphi(x) - \varphi(x_0)}{\psi(x) - \psi(x_0)} = \frac{\varphi'(\xi)}{\psi'(\xi)} \quad (\xi \text{ between } x \text{ and } x_0),$$

whilst, on the other hand,

$$\frac{\varphi(x) - \varphi(x_0)}{\psi(x) - \psi(x_0)} = \frac{\varphi(x)}{\psi(x)} \cdot \frac{1 - \dfrac{\varphi(x_0)}{\varphi(x)}}{1 - \dfrac{\psi(x_0)}{\psi(x)}}.$$

We remark that it follows directly from (8) that $\varphi(x)$ and $\psi(x)$ differ from zero for x near a.

Comparison of these two equations gives us, after re-arrangement:

$$\frac{\varphi(x)}{\psi(x)} = \frac{\varphi'(\xi)}{\psi'(\xi)} \cdot \frac{1 - \dfrac{\psi(x_0)}{\psi(x)}}{1 - \dfrac{\varphi(x_0)}{\varphi(x)}}, \tag{10}$$

where ξ lies between x and x_0, and hence between a and x_0. We take x_0 sufficiently near a, so that, by (9), the first factor on the right of (10) differs from b by an arbitrarily small amount, for any choice of x between x_0 and a. Having thus fixed x_0, we let x approach a. The second factor on the right of (10) now tends to unity, by (8), and hence we can say that the left-hand side of (10) differs from b by an arbitrarily small amount for values of x near a, i.e.

$$\lim_{x \to a} \frac{\varphi(x)}{\psi(x)} = b.$$

It follows from the theorem just proved that l'Hôpital's rule can be used for evaluating an indeterminacy of the form ∞/∞.

We note some further indeterminate forms. We take the product $\varphi(x)\psi(x)$, and let

$$\lim_{x \to a} \varphi(x) = 0 \text{ and } \lim_{x \to a} \psi(x) = \infty.$$

This gives the indeterminate form $0.\infty$. It is easily transformed to the form $0/0$ or ∞/∞:

$$\varphi(x)\,\psi(x) = \frac{\varphi(x)}{\dfrac{1}{\psi(x)}} = \frac{\psi(x)}{\dfrac{1}{\varphi(x)}}.$$

We consider finally the expression $\varphi(x)^{\psi(x)}$ and let

$$\lim_{x \to a} \varphi(x) = 1 \text{ and } \lim_{x \to a} \psi(x) = \infty.$$

This gives the indeterminate form 1^∞. We take the logarithm of the above expression:

$$\log\,[\varphi(x)^{\psi(x)}] = \psi(x) \log \varphi(x)\,,$$

giving the indeterminate form $0 \cdot \infty$. By evaluating this indeterminacy, i.e. finding the limit of the logarithm of the given expression, we can discover the limit of the expression itself. The indeterminate forms ∞^0 and 0^0 are similarly evaluated.

We now take some examples:

1.
$$\lim_{x\to+\infty} \frac{e^x}{x} = \lim_{x\to+\infty} \frac{e^x}{1} = +\infty,$$

$$\lim_{x\to+\infty} \frac{e^x}{x^2} = \lim_{x\to+\infty} \frac{e^x}{2x} = \lim_{x\to+\infty} \frac{e^x}{2} = +\infty.$$

It can similarly be shown that e^x/x^n tends to infinity as x tends to $+\infty$, for any positive n, i.e. *the exponential function* e^x *increases more rapidly than any positive power of* x, on indefinite increase of x.

2.
$$\lim_{x\to+\infty} \frac{\log x}{x^n} = \lim_{x\to+\infty} \frac{\frac{1}{x}}{nx^{n-1}} = \lim_{x\to+\infty} \frac{1}{nx^n} = 0 \quad (n > 0),$$

i.e. $\log x$ *increases more slowly than any positive power of* x.

3.
$$\lim_{x\to+0} x^n \log x = \lim_{x\to+0} \frac{\log x}{\frac{1}{x^n}} = \lim_{x\to+0} \frac{\frac{1}{x}}{\frac{-n}{x^{n+1}}} =$$

$$= -\lim_{x\to+0} \frac{x^n}{n} = 0 \qquad (n > 0).$$

4. We find the limit of x^x for $x \to 0$. The logarithm of the expression gives the indeterminate form $0 \cdot \infty$. This indeterminate form has limit 0, by Example 3, and hence:

$$\lim_{x\to+0} x^x = 1.$$

5. We find the limit of the ratio:

$$\lim_{x\to\infty} \frac{x + \sin x}{x}.$$

Numerator and denominator both tend to infinity. Using the rule for replacing the ratio of the functions by the ratio of their derivatives, we get:

$$\lim_{x\to\infty} \frac{1 + \cos x}{1}.$$

But $1 + \cos x$ tends to no definite limit on indefinite increase of x, since $\cos x$ always oscillates between 1 and -1; yet it is easy to see that the given ratio itself tends to a limit:

$$\lim_{x\to\infty} \frac{x + \sin x}{x} = \lim_{x\to\infty} \left(1 + \frac{\sin x}{x}\right) = 1.$$

The indeterminate form has a limit here, yet the rule for finding it fails. This result does not contradict the theorem, since the theorem states only that, if the ratio of the derivatives tends to a limit, the ratio of the functions tends to the same limit: the theorem does not state the converse.

6. We also note the indeterminate form $(\infty \pm \infty)$. It usually leads to the form $0/0$. For example:

$$\lim_{x\to 0}\left(\frac{1}{\sin x}-\frac{1}{x+x^2}\right)=\lim_{x\to 0}\frac{x+x^2-\sin x}{(x+x^2)\sin x}.$$

The last expression consists of the indeterminate form $0/0$. We evaluate this by the method given above:

$$\lim_{x\to 0}\left(\frac{1}{\sin x}-\frac{1}{x+x^2}\right)=1.$$

§ 6. Functions of two variables

67. Basic concepts. We have so far considered functions of a single independent variable. We now consider a function of two independent variables

$$u=f(x, y).$$

Values must be attributed to the independent variables: $x = x_0$, $y = y_0$, in order to define particular values of this function. Every such pair of values of x and y defines a corresponding point M_0 in the plane of coordinates, with coordinates (x_0, y_0), and we can speak about the value of the function at the point $M_0(x_0, y_0)$ of the plane, instead of the value for $x = x_0$, $y = y_0$. The function can be defined *over the whole plane* or only in some part of it, *in a certain domain.* If $f(x, y)$ is an integral polynomial in x, y, e.g.,

$$u=f(x, y)=x^2+xy+y^2-2x+3y+7,$$

the function can be considered to be defined by its formula over the whole plane. The formula:

$$u=\sqrt{1-(x^2+y^2)}$$

defines the function inside the circle $x^2+y^2=1$ with centre at the origin and radius unity, and also on the circle itself, where $u = 0$. A similar interval on the plane is given by the domain, defined by the inequalities $a \leqslant x \leqslant b$; $c \leqslant y \leqslant d$. This is a rectangle with sides parallel to the axes, the boundaries of the rectangle being also included in the domain. The inequalities $a < x < b$, $c < y < d$ define only interior points of the rectangle. If the boundaries of the domain are included in it, the domain is said to be *closed*. If the boundaries are not included in the domain, it is said to be *open* (cf. [4]). We shall

define the concept of limit for functions of two variables (cf. [32]). We suppose that the function is defined at the point $M_0(x_0, y_0)$ and at all points $M(x, y)$ sufficiently near M_0.

DEFINITION. *If a function $f(x, y)$ has the same limit A, whatever the rule by which the point $M(x, y)$ tends to $M_0(x_0, y_0)$, we write:*

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} f(x, y) = A$$

or

$$\lim_{M \to M_0} f(x, y) = A.$$

This definition is equivalent to the following: for any given positive ε there exists a positive η such that:

$$|f(x, y) - A| < \varepsilon, \quad \text{if} \quad |x - x_0| < \eta \quad \text{and} \quad |y - y_0| < \eta,$$

the pair of values $x = x_0$ and $y = y_0$ being excluded. The values $x = x_0$, $y \neq y_0$ or $x \neq x_0$, $y = y_0$ remain possible. If M_0 lies on the boundary of the domain in which $f(x, y)$ is defined, M must tend to M_0 in such a way as to belong to this domain.

The definition of continuity follows naturally (cf. [32]).

DEFINITION. *A function $f(x, y)$ is said to be continuous at the point $M_0(x_0, y_0)$, if:*

$$\lim_{\substack{x \to x_0 \\ y \to y_0}} f(x, y) = f(x_0, y_0) \text{ or } \lim_{M \to M_0} f(x, y) = f(x_0, y_0).$$

A function is said to be continuous in a certain domain if it is continuous at every point of this domain.

The function $w = \sqrt{1 - x^2 - y^2}$, for example, is continuous inside the circle in which it is defined. It can also be said of it, that it remains continuous if we associate the boundary with the circular domain, i.e. the circumference where $w = 0$.

Function $f(x, y)$, continuous in a certain domain, including its boundary, has the following two properties, analogous to those of a function of a single independent variable, continuous in a certain interval [35]:

(1) there is at least one point in the domain or on the boundary where it takes its greatest (least) value, relative to its remaining values in the domain, including the boundary;

(2) it is uniformly continuous in the domain, including the boundary, i.e. for any given positive ε there exists one positive η for all the domain, such that

$$|f(x_2, y_2) - f(x_1, y_1)| < \varepsilon \quad \text{if} \quad |x_2 - x_1| \quad \text{and} \quad |y_2 - y_1| < \eta,$$

where (x_1, y_1) and (x_2, y_2) are points of the domain.

We note a consequence of the definition of the continuity of a function. If $f(x, y)$ is continuous at the point (a, b), and if we put $y = b$, the function $f(x, b)$ of a single independent variable x is continuous at $x = a$. Similarly, $f(a, y)$ is continuous for $y = b$.

68. The partial derivatives and total differential of a function of two independent variables. We suppose that y is constant, and only x varies, in the function $u = f(x, y)$; u becomes a function of the single variable x and its increment and derivative can be found. Let $\Delta_x u$ denote the increment of u, when y is constant, and x has the increment Δx:

$$\Delta_x u = f(x + \Delta x, y) - f(x, y).$$

We get the derivative by finding the limit:

$$\lim_{\Delta x \to 0} \frac{\Delta_x u}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x}.$$

This derivative, obtained with y constant, is called the *partial derivative of u with respect to x*, and is denoted by:

$$\frac{\partial f(x, y)}{\partial x} \quad \text{or} \quad f'_x(x, y) \quad \text{or} \quad \frac{\partial u}{\partial x}.$$

We note that $\partial u/\partial x$ is not to be taken as a fraction, being purely the symbol for the partial derivative. If $f(x, y)$ has a partial derivative with respect to x, it is a continuous function of x with y fixed.

The increment $\Delta_y u$ is defined in exactly the same way, as is the partial derivative of u with respect to y, found by assuming that x does not vary:

$$\frac{\partial f(x, y)}{\partial y} \quad \text{or} \quad f'_y(x, y) \quad \text{or} \quad \frac{\partial u}{\partial y} = \lim_{\Delta y \to 0} \frac{\Delta_y u}{\Delta y} =$$

$$= \lim_{\Delta y \to 0} \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y}.$$

If, for example,

$$u = x^2 + y^2, \quad \text{then} \frac{\partial u}{\partial x} = 2x, \qquad \frac{\partial u}{\partial y} = 2y.$$

We take Clapeyron's equation:

$$pv = RT.$$

One of the magnitudes p, v and T can be defined with the aid of this equation in terms of the other two, these latter two being considered as independent variables. We get the table:

Independent variables	T, p	T, v	p, v
Function	$v = \frac{RT}{p}$	$p = \frac{RT}{v}$	$T = \frac{pv}{R}$
Partial derivatives	$\frac{\partial v}{\partial T} = \frac{R}{p}$; $\frac{\partial v}{\partial p} = -\frac{RT}{p^2}$	$\frac{\partial p}{\partial T} = \frac{R}{v}$; $\frac{\partial p}{\partial v} = -\frac{RT}{v^2}$	$\frac{\partial T}{\partial p} = \frac{v}{R}$; $\frac{\partial T}{\partial p} = \frac{p}{R}$

The following relationship is thus obtained:

$$\frac{\partial v}{\partial T} \cdot \frac{\partial T}{\partial p} \cdot \frac{\partial p}{\partial V} = -1.$$

If we were to reduce the left-hand side of this equation as a fraction, we should get $(+1)$ instead of (-1). But the partial derivatives in this equation are worked out on different assumptions: $\partial v/\partial T$, on the assumption that p is constant; $\partial T/\partial p$, on the assumption that v is constant; and $\partial p/\partial v$ with T constant. The reduction mentioned is thus not permissible.

We let Δu denote the total increment of the function, obtained with simultaneous variation of x and y:

$$\Delta u = f(x + \Delta x, y + \Delta y) - f(x, y).$$

Adding and subtracting $f(x, y + \Delta y)$, we can write:

$$\Delta u = [f(x + \Delta x, y + \Delta y) - f(x, y + \Delta y)] + \\ + [f(x, y + \Delta y) - f(x, y)].$$

The first square bracket gives us the increment of u with variable y fixed at $(y + \Delta y)$, whilst the second square bracket gives us the increment of the same function with x fixed. We now assume that $f(x, y)$ has partial derivatives over a domain that includes (x, y) as

an interior point, and we apply Lagrange's formula to each of these increments, as can be done, since only one independent variable changes in each case; we get:

$$\Delta u = f'_x(x + \theta\Delta x, y + \Delta y)\Delta x + f'_y(x, y + \theta_1 \Delta y)\Delta y ,$$

where θ and θ_1 lie between zero and unity. Assuming the continuity of the partial derivative $\partial u/\partial x$ and $\partial u/\partial y$, we can say that the coefficient of Δx tends to $f'_x(x, y)$, and the coefficient of Δy to $f'_x(x, y)$, as Δx and Δy tend to zero; hence,

$$\Delta u = [f'_x(x, y) + \varepsilon]\Delta x + [f'_x(x, y) + \varepsilon_1]\Delta y$$

or

$$\Delta u = f'_x(x, y)\Delta x + f'_y(x, y)\Delta y + \varepsilon\Delta x + \varepsilon_1\Delta y , \tag{1}$$

where ε and ε_1 are infinitesimals respectively with Δx and Δy. This is analogous to the formula:

$$\Delta y = y'\Delta x + \varepsilon\Delta x ,$$

that we obtained for a function of a single independent variable [50]. The products $\varepsilon\Delta x$ and $\varepsilon_1\Delta y$ are infinitesimals of higher order as regards Δx and Δy respectively.

We recall that we based the above argument on the assumption not only of the existence, but also of the continuity, of the partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ over a certain domain, including (x, y) as an interior point.

The sum of the first two terms appearing on the right of equation (1) is called *the total differential* du *of function* u. The arbitrary increments Δx and Δy of the independent variables coincide with their differentials dx and dy, as in the case of a single independent variable, so that

$$\mathrm{d}u = f'_x(x, y)\mathrm{d}x + f'_y(x, y)\mathrm{d}y ,$$

or

$$\mathrm{d}u = \frac{\partial u}{\partial x}\,\mathrm{d}x + \frac{\partial u}{\partial y}\,\mathrm{d}y . \tag{2}$$

We can say, in view of the property mentioned above of the products $\varepsilon\Delta x$ and $\varepsilon_1\Delta y$, that for small values of Δx and Δy *the total differential* du *gives the approximate magnitude of the total increment of the function* Δu. Evidently, on the other hand, the products $\partial u/\partial x\,\mathrm{d}x$ and $\partial u/\partial y\,\mathrm{d}y$ give the approximate magnitudes of the increments $\Delta_x u$ and $\Delta_y u$,

and thus, *for small increments of the independent variables, the total increment of the function is approximately equal to the sum of its partial increments :*

$$\Delta u \sim \mathrm{d}u \sim \Delta_x u + \Delta_y u .$$

Equation (2) expresses an extremely important property of functions of several independent variables, which may be called the "principle of combination of small operations". Its essence lies in the fact that the combined effect of several small operations, say Δx and Δy, can be replaced with sufficient accuracy by the sum of the effects of each separate small operation.

69. Derivatives of functions of a function and of implicit functions. We now suppose that the function $u = f(x, y)$ depends on a single independent variable t through the medium of x and y, i.e. we take x and y as functions of the independent variable t instead of being themselves independent variables. We define the derivative $\partial u/\partial t$ of u with respect to t.

If t receives the increment Δt, the functions x and y receive the corresponding increments Δx and Δy, and u receives the increment Δu:

$$\Delta u = f(x + \Delta x, y + \Delta y) - f(x, y) .$$

We saw in [68] that the increment can be written in the form:

$$\Delta u = f'_x(x + \theta \Delta x, y + \Delta y)\Delta x + f'_y(x, y + \theta_1 \Delta y)\Delta y .$$

We divide both sides of this equation by Δt:

$$\frac{\Delta u}{\Delta t} = f'_x(x + \theta\Delta x, y + \Delta y)\frac{\Delta x}{\Delta t} + f'_y(x, y + \theta_1 \Delta y)\frac{\Delta y}{\Delta t} .$$

We have assumed that x and y have derivatives with respect to t, and thus are certainly continuous functions of t. Hence Δx and Δy tend to zero as Δt tends to zero, and, assuming the continuity of $\partial u/\partial x$ and $\partial u/\partial y$, the above equation gives us in the limit:

$$\frac{\mathrm{d}u}{\mathrm{d}t} = f'_x(x, y)\frac{\mathrm{d}x}{\mathrm{d}t} + f'_y(x, y)\frac{\mathrm{d}y}{\mathrm{d}t} . \tag{3}$$

This equation gives *the rule for differentiating a function of a function, in the case of a function of more than one variable.*

We suppose that, in particular, x has the role of the independent variable t, i.e. that the function $u = f(x, y)$ depends on the independent variable x both directly, and via the medium of y, a function of x. Noting that $dx/dx = 1$, we get from the fundamental equation (3):

$$\frac{du}{dx} = f'_x(x, y) + f'_y(x, y)\frac{dy}{dx}. \tag{4}$$

The derivative du/dx is called the total derivative of u with respect to x, as distinct from the *partial* derivative $f'_x(x, y)$.

The rule for differentiation just given can be used for finding *the derivative of an implicit function.* Let

$$F(x, y) = 0 \tag{5}$$

define y as an implicit function of x, having the derivative

$$y' = \varphi'(x).$$

Substituting $y = \varphi(x)$ in (5) would necessarily give us the identity $0 = 0$, since $y = \varphi(x)$ is a solution of (5). We thus see that the constant zero can be regarded as a function of a function of x, dependent on x both directly and through the medium of $y = \varphi(x)$.

The derivative with respect to x of this constant must be zero; applying rule (4), we get:

$$F'_x(x, y) + F'_y(x, y)y' = 0,$$

whence

$$y' = -\frac{F'_x(x, y)}{F'_y(x, y)}.$$

Both x and y can enter into the expression thus obtained for y', and if an expression has to be found for y' only in terms of the independent variable x, it still comes to a question of solving (5) with respect to y.

§ 7. Some geometrical applications of the differential calculus

70. The differential of arc. It will be shown in the integral calculus how to find the length of an arc of a curve; the expression for the differential of the length of arc will then be investigated, and it will

be shown that the ratio of the length of the chord to the length of the arc subtending it tends to unity, when the arc contracts indefinitely to a point.

Let $y = f(x)$ be a given curve, and let the length of arc be measured on it from some fixed point A in a specified direction (Fig. 73). Let s be the length of arc from A to a variable point M. The magnitude s, like the ordinate y, will be a function of the abscissa x of M. If the direction of AM coincides with the specified direction of the curve, $s > 0$, whilst in the opposite case, $s < 0$. Let $M(x, y)$ and $N(x + \Delta x, y + \Delta y)$ be two points of the curve, and Δs the difference of the lengths of arc AN and AM, i.e. the increment in length of arc in passing from M to N. The absolute value of Δs is the length of arc MN taken with the plus sign. We have from the right-angled triangle:

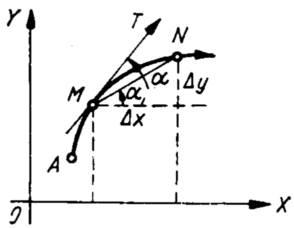

FIG. 73

$$(\overline{MN})^2 = \Delta x^2 + \Delta y^2 ,$$

whence

$$\frac{(\overline{MN})^2}{\Delta x^2} = 1 + \left(\frac{\Delta y}{\Delta x}\right)^2$$

or

$$\left(\frac{\overline{MN}}{\Delta s}\right)^2 \left(\frac{\Delta s}{\Delta x}\right)^2 = 1 + \left(\frac{\Delta y}{\Delta x}\right)^2 .$$

Passing to the limit, and noting that $(\overline{MN}/\Delta s)^2 \to 1$ from what was remarked above, we get

$$\left(\frac{\mathrm{d}s}{\mathrm{d}x}\right)^2 = 1 + \left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)^2$$

or

$$\frac{\mathrm{d}s}{\mathrm{d}x} = \pm \sqrt{1 + y'^2} . \tag{1}$$

We must take the positive sign if s increases as x increases, and the negative sign if s decreases with increasing x. We suppose the former case for clarity (illustrated in Fig. 73). It follows from (1) that

$$\mathrm{d}s = \sqrt{1 + y'^2}\, \mathrm{d}x$$

or

$$\mathrm{d}s = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2} . \tag{2}$$

The length s of arc AM is a natural parameter for defining the position of a point M on a curve. We can take s as independent variable, the coordinates (x, y) of a point M then being functions of s:

$$x = \varphi(s); \qquad y = \psi(s).$$

A more detailed discussion of the "parameters of a curve" is given in [74]. We now explain the geometrical meaning of the derivatives of x and y with respect to s.

We take the point N so that the direction of arc MN coincides with the specified direction of the curve, i. e. $\Delta s > 0$. The direction of the chord $\overline{MN}$ gives in the limit, as N tends to M, a specified direction of the tangent at M. We take this as the positive direction of the tangent. It is associated with the direction specified for the curve itself.

Let α_1 be the angle between the direction of $\overline{MN}$ and the positive direction of axis OX. The increment Δx of the abscissa x is the projection of $\overline{MN}$ on OX, and hence:

$$\Delta x = \overline{MN} \cos \alpha_1,$$

$$(\overline{MN} = \sqrt{\Delta x^2 + \Delta y^2}),$$

$\overline{MN}$ being reckoned positive in this equation. On dividing both sides of the equation by the length of arc MN, equal to Δs, we get:

$$\frac{\Delta x}{\Delta s} = \frac{\sqrt{\Delta x^2 + \Delta y^2}}{\Delta s} \cos \alpha_1.$$

By hypothesis, $\Delta s > 0$, and hence the ratio $\sqrt{\Delta x^2 + \Delta y^2}/\Delta s$ tends to $(+1)$ as N tends to M, whilst α_1 tends to α, the angle between the positive direction of the tangent $\overline{MT}$ and the positive direction of OX. The equation just given becomes in the limit:

$$\cos \alpha = \frac{dx}{ds}. \tag{3}$$

Similarly, by projecting $\overline{MN}$ on OY, we get:

$$\sin \alpha = \frac{dy}{ds}. \tag{4}$$

71. Concavity, convexity and curvature. Curves, convex and concave towards positive ordinates, are shown in Figs. 74 and 75.

The same curve $y = f(x)$ can, of course, have convex and concave portions (Fig. 76). *Points separating convex from concave portions of a curve are called points of inflexion.*

If we move along the curve in the direction of increasing x, and observe the variation of α, the angle formed by the tangent with

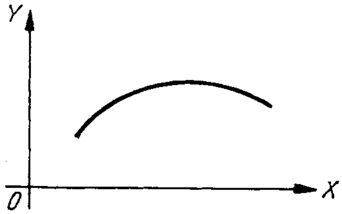

FIG. 74

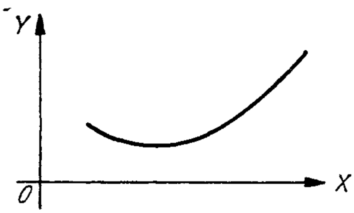

FIG. 75

the positive direction of OX, we see that (Fig. 76) *α is decreasing in convex, and increasing in concave, portions.* It follows that $\tan \alpha$, i.e. the derivative $f'(x)$, will undergo the same variation, since $\tan \alpha$ increases (decreases) with increasing (decreasing) α. But the interval in which $f'(x)$ is decreasing is the interval where its derivative is negative, i.e. $f''(x) < 0$; and similarly, the interval of increase of $f'(x)$ is that where $f''(x) > 0$. We thus have the theorem:

The portions of a curve are convex towards positive ordinates where $f''(x) < 0$, being concave where $f''(x) > 0$. Points of inflexion are the points where $f''(x)$ changes sign.

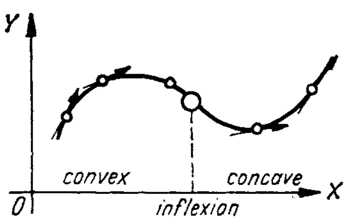

FIG. 76

By using arguments analogous to the earlier ones of [58], we obtain from this theorem a rule for finding the points of inflexion of a curve: *to find the points of inflexion of a curve, the values of x must be found for which $f''(x)$ vanishes or does not exist, and the variation in sign of $f''(x)$ must be investigated on passage through these values, using the following table :*

	Point of inflexion		Not a point of inflex	
$f''(x)$	$+-$	$-+$	$--$	$++$
	conc. conv.	conv. conc.	convex	concave

The most natural way of representing the bending of a curve is by following the variation of a, the angle made by the tangent with OX, as we move along the curve. Given two arcs of the same length Δs, the more curved arc will be that for which the tangent moves through a greater angle, i.e. for which the increment Δa is greater. This remark leads us to the concepts of the mean curvature of Δs and of the curvature at a given point: *the mean curvature of an arc Δs is defined as the absolute value of the ratio of Δa, the angle between the tangents at the ends of the arc, to the length of arc Δs. The limit of this ratio as Δs tends to zero is called the curvature of the curve at the given point* (Fig. 77).

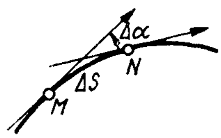

FIG. 77

We thus have for the curvature C:

$$C = \left|\frac{\mathrm{d}a}{\mathrm{d}s}\right|.$$

But $\tan a$ is the first derivative y', i. e.

$$a = \operatorname{arc}\tan y',$$

whence, differentiating the function of a function $\operatorname{arc}\tan y'$ with respect to x:

$$\mathrm{d}a = \frac{y''}{1 + y'^2}\,\mathrm{d}x.$$

As shown above,

$$\mathrm{d}s = \pm\sqrt{1 + y'^2}\,\mathrm{d}x.$$

Dividing $\mathrm{d}a$ by $\mathrm{d}s$, we get a final expression for the curvature:

$$C = \pm\frac{y''}{(1 + y'^2)^{3/2}}. \tag{5}$$

The minus sign is taken in convex parts, and the plus sign in concave parts, so as to give C a positive value.

No curvature exists at points of a curve where the derivatives y' or y'' do not exist. A curve resembles a straight line in the neighbourhood of points where y'', and hence the curvature, vanishes; this will happen, for instance, near points of inflexion.

Suppose the coordinates x, y of a point of a curve are given in terms of the length of arc s. Here, as we have seen:

$$\cos a = \frac{\mathrm{d}x}{\mathrm{d}s}, \quad \sin a = \frac{\mathrm{d}y}{\mathrm{d}s}.$$

Angle a is also a function of s, and on differentiating the above equations with respect to s, we get:

$$-\sin a\frac{da}{ds}=\frac{d^2x}{ds^2},\qquad \cos a\frac{da}{ds}=\frac{d^2y}{ds^2}.$$

Squaring both sides of these equations and adding, we have:

$$\left(\frac{da}{ds}\right)^2=\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2 \text{ or } C^2=\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2,$$

whence:

$$C=\sqrt{\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2}.$$

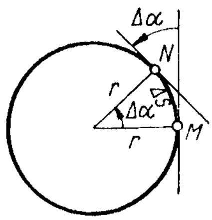

FIG. 78

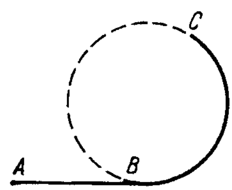

FIG. 79

The reciprocal of the curvature, $1/C$, is called the *radius of curvature*. By (5), we have the following expression for the radius of curvature R:

$$R=\left|\frac{ds}{da}\right|=\pm\frac{(1+y'^2)^{3/2}}{y''} \tag{6}$$

or

$$R=\frac{1}{\sqrt{\left(\frac{d^2x}{ds^2}\right)^2+\left(\frac{d^2y}{ds^2}\right)^2}},$$

taking the positive value of the square root.

In the case of a straight line, y is a polynomial of the first degree in x, and hence y'' is identically zero, i.e. the curvature is zero everywhere on the line, the radius of curvature being infinity.

We evidently have for a circle of radius r (Fig. 78):

$$\Delta s=r\Delta a \quad\text{and}\quad R=\lim\frac{\Delta s}{\Delta a}=r,$$

i.e. the radius of curvature is constant for the entire circle. We shall see later that the circle alone has this property.

We remark that the variation of the radius of curvature is by no means as easily seen as that of the tangent. We take the curve made up of an arc BC of a circle and a section AB of the tangent to the circle at B (Fig. 79). The radius of curvature is infinity for the portion AB, whilst for the portion

BC it is equal to the radius of the circle r; it thus suffers a break in continuity at B, whereas the direction of the tangent varies continuously here. This accounts for the jolting of railway carriages at bends. Assuming a carriage travelling with fixed speed v, we know from mechanics that a force is exerted along the normal to the trajectory, equal to $m\,v^2/R$, where m is the mass of the moving body, and R is the radius of curvature of the trajectory. Hence the force suffers a break in continuity at points where the radius of curvature suffers a break in continuity, which explains the jolts.

72. Asymptotes. We now turn to considering curves with *infinite branches*, where one or both of the coordinates x and y increase indefinitely. The hyperbola and parabola are examples of such curves.

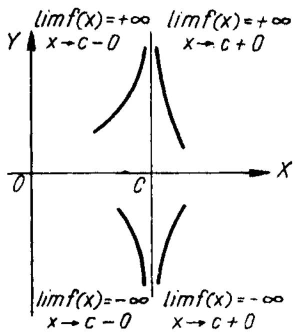

FIG. 80

A straight line is referred to as an asymptote of a curve with an infinite branch when the distance of points of the curve from the line tends to zero on indefinite displacement along the infinite branch.

We first show how to find the asymptotes of a curve parallel to axis OY. The equation of an asymptote of this sort must have the form:

$$x = c\,,$$

where c is a constant, and x must tend to c, whilst y tends to infinity, on moving along the corresponding infinite branch (Fig. 80). We thus get the following rule:

All the asymptotes parallel to OY of the curve

$$y = f(x)$$

can be found by finding all the values $x = c$, on approach to which $f(x)$ tends to infinity.

To find the position of the curve relative to the asymptote, the sign of $f(x)$ must be determined as x tends to c on the left and right.

We pass to finding the asymptotes, not parallel to OY. The equation of the asymptote must now have the form:

$$\eta = a\,\xi + b\,,$$

where ξ, η are the current coordinates of the asymptote, as distinct from x, y, the current coordinates of the curve.

Let ω be the angle that the asymptote forms with the positive direction of OX, let $\overline{MK}$ be the distance of a point of the curve from the asymptote, and $\overline{MK_1}$ be the difference between the ordinates of the curve and asymptote for the same abscissa x (Fig. 81). We have from the right-angled triangle:

$$\overline{MK_1} = \frac{\overline{MK}}{\cos\omega} \qquad \left(\omega \neq \frac{1}{2}\pi\right),$$

and hence we can replace the condition:

$$\lim_{x\to\infty} \overline{MK} = 0$$

by the condition

$$\lim_{x\to\infty} \overline{MK_1} = 0. \tag{7}$$

For an asymptote not parallel to OY, x tends to infinity on moving along the infinite branch. Recalling that $\overline{MK_1}$ is the difference between the ordinates of the curve and asymptote for the same abscissa, we can rewrite condition (7) as:

$$\lim_{x\to\infty} [f(x) - ax - b] = 0, \tag{8}$$

where the values of a and b have to be found.

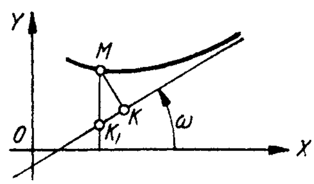

FIG. 81

We can rewrite (8) in the form:

$$\lim_{x\to\infty} x\left[\frac{f(x)}{x} - a - \frac{b}{x}\right] = 0;$$

the first factor x tends to infinity, so that the expression in square brackets must tend to zero:

$$\lim_{x\to\infty}\left[\frac{f(x)}{x} - a - \frac{b}{x}\right] = \lim_{x\to\infty}\frac{f(x)}{x} - a = 0,$$

i.e.

$$a = \lim_{x\to\infty}\frac{f(x)}{x}.$$

Having found a, we obtain b from the basic condition (8), which can be written as:

$$b = \lim_{x\to\infty}[f(x) - ax].$$

Thus, *a necessary and sufficient condition that the curve:*

$$y = f(x)$$

has an asymptote not parallel to OY is that, with x increasing indefinitely on movement along the infinite branch, the limits exist :

$$a = \lim_{x \to \infty} \frac{f(x)}{x}, \quad b = \lim_{x \to \infty} [f(x) - ax],$$

the equation of the asymptote then being :

$$\eta = a\,\xi + b.$$

To find the position of the curve relative to its asymptote, the cases of x tending to $(+\infty)$ and to $(-\infty)$ have to be worked out separately, and the sign of the difference,

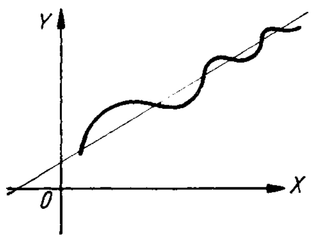

FIG. 82

$$f(x) - (ax + b),$$

determined in each case. If the sign is positive, the curve is situated above the asymptote, and if negative, below the asymptote. If the difference does not keep the same sign on indefinite increase of x, the curve will oscillate about the asymptote (Fig. 82).

73. Curve-tracing. We now give a fuller indication than in [59] of the series of operations to be carried out in tracing the curve

$$y = f(x).$$

We must:

(a) define the interval of variation of the independent variable x;

(b) find the points of intersection of the curve with the coordinate axes;

(c) find the peaks of the curve;

(d) find the convexities, concavities, and points of inflexion of the curve;

(e) find the asymptotes of the curve;

(f) examine the symmetry of the curve relative to the coordinate axes, if such symmetry exists.

The curve can be traced more accurately if an extra series of points on it are plotted. The coordinates of these points can be found from the equation of the curve.

1. We trace the curve

$$y = \frac{(x-3)^2}{4(x-1)}.$$

(a) x can vary in the interval $(-\infty, +\infty)$.

(b) Putting $x = 0$, we get $y = -9/4$; putting $y = 0$, we get $x = 3$, i.e. the curve intercepts the axes at the points $(0, -9/4)$ and $(3,0)$.

(c) We find the first and second derivatives:

$$f'(x) = \frac{(x-3)(x+1)}{4(x-1)^2}, \quad f''(x) = \frac{2}{(x-1)^3}.$$

We apply the usual rule to find the peaks: $(3,0)$ is a minimum, $(-1, -2)$ is a maximum.

(d) It is clear from the expression for the second derivative that it is positive for $x > 1$, and negative for $x < 1$, i.e. the curve is concave in the interval $(1, \infty)$, and convex in $(-\infty, 1)$. There is no point of inflexion, since $f''(x)$ changes sign only at $x = 1$, where the curve has an asymptote parallel to OY, as we shall see next.

(e) y becomes infinite at $x = 1$, and the curve has an asymptote:

$$x = 1.$$

We now look for asymptotes, not parallel to OY:

$$a = \lim_{x\to\infty} \frac{(x-3)^2}{4(x-1)x} = \lim_{x\to\infty} \frac{\left(1 - \frac{3}{x}\right)^2}{4\left(1 - \frac{1}{x}\right)} = \frac{1}{4},$$

$$b = \lim_{x\to\infty} \frac{(x-3)^2}{4(x-1)} - \frac{x}{4} = \lim_{x\to\infty} \frac{-5x+9}{4(x-1)} =$$

$$= \lim_{x\to\infty} \frac{-5 + \frac{9}{x}}{4\left(1 - \frac{1}{x}\right)} = -\frac{5}{4},$$

The asymptote is thus:

$$y = \frac{1}{4}x - \frac{5}{4}.$$

We propose to the reader the investigation of the position of the curve relative to the asymptotes.

(f) Symmetry does not exist.

Transferring all the data obtained to the figure, we get the curve of Fig. 83.

2. We investigate the curves:

$$y = c(a^2 - x^2)(5a^2 - x^2) \quad (c < 0)$$

$$y_1 = c(a^2 - x^2)^2,$$

which give the shape of a heavy beam, bending under its own weight, the first curve referring to the case when the ends are freely supported, and the second to the case when the ends are constrained. The total length of the beam is $2a$, the origin is at the centre of the beam, and axis OY is directed vertically upwards.

(a) The variation of x evidently only interests us in the interval $(-a, +a)$

(b) Putting $x = 0$, we get $y = 5ca^4$ and $y_1 = ca^4$, i.e. the bending at the centre of the beam is five times greater in the first case than in the second. For $x = \pm a$, $y = y_1 = 0$, corresponding to the ends of the beam.

(c) We find the derivatives:

$$y' = -4cx(3a^2 - x^2), \qquad y'' = -12c(a^2 - x^2)\,,$$

$$y_1' = -4cx(a^2 - x^2)\,, \qquad y_1'' = -4c(a^2 - 3x^2)\,.$$

There will be a minimum at $x = 0$ in the interval $(-a, +a)$ in both cases, corresponding to the bending of the centre of the beam, mentioned above.

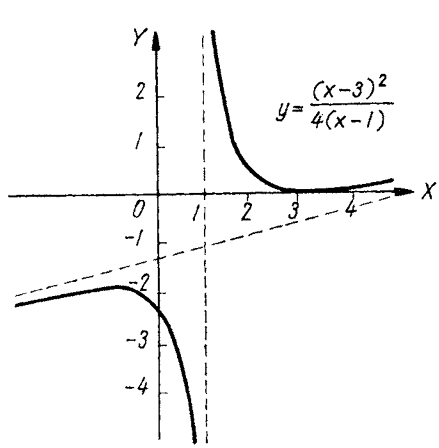

FIG. 83

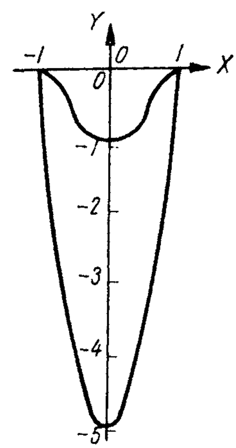

FIG. 84

(d) In the first case, $y'' > 0$ in the interval $(-a, +a)$, i.e. the entire beam is concave upwards. In the second case, y_1'' vanishes at $x = -a/\sqrt{3}$; its sign changes here, and the corresponding points are points of inflexion of the beam.

(e) There are no infinite branches.

(f) Both equations remain unchanged on replacing x by $(-x)$, i.e. both curves are symmetrical about OY.

The two curves are shown in Fig. 84. We have taken the cases $a = 1$, $c = -1$, for simplicity; the length of the beam is considerably greater than its bending, in practice, i.e. a is considerably greater than c, so that the curve of bending looks rather different (how?).

We suggest that the points of inflexion of the curve:

$$y = e^{-x^2}$$

might be found by the reader, and a comparison made with the graph of Fig. 60.

74. The parameters of a curve. Given the properties of a geometrical locus, it is not always convenient or possible to find its equation by expressing its properties directly in terms of an equation connecting the current coordinates x and y. It is useful in this case to introduce a third, auxiliary variable, in terms of which the abscissa x and ordinate y of any given point of the locus can be separately expressed.

The combination of two equations obtained in this way:

$$x = \varphi(t)\,, \qquad y = \psi(t) \tag{9}$$

can also be used for plotting and investigating a curve, since each value of t defines the position of a corresponding point of the curve.

This method is referred to as *parametric representation of a curve*, the auxiliary variable t being a parameter. To obtain the equation of the curve in the usual (explicit or implicit) form as a relationship between x and y, the parameter t must be *eliminated* from equations (9), as might possibly be done by solving one of the equations with respect to t, and substituting the result in the other.

Curves given by parameters are especially met with in mechanics, as when finding the trajectory of motion of a point, the position of which depends on time t, so that its coordinates are functions of t. The trajectory is given by a parameter, when these functions are known.

For instance, the equation of a circle with centre at (x_0, y_0) and radius r is given in terms of a parameter as:

$$x = x_0 + r\cos t; \quad y = y_0 + r\sin t. \tag{10}$$

We rewrite these equations as:

$$x - x_0 = r\cos t; \quad y - y_0 = r\sin t.$$

We eliminate the parameter t by squaring both sides and adding, which gives the ordinary equation of a circle:

$$(x - x_0)^2 + (y - y_0)^2 = r^2.$$

Similarly, it is immediately obvious that

$$y = a\cos t; \quad y = b\sin t \tag{11}$$

are the parametric equations of the ellipse:

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1.$$

Let y be defined parametrically by formula (9) as a function of x. An increment Δt of the parameter produces corresponding increments Δx and Δy, and by dividing numerator and denominator of Δx by Δt, we find the following expression for the derivative of y with respect to x:

$$y'_x = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\frac{\Delta y}{\Delta t}}{\frac{\Delta x}{\Delta t}} = \frac{\psi'(t)}{\varphi'(t)}$$

or

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\psi'(t)}{\varphi'(t)}. \tag{12}$$

We find the second derivative of y with respect to x:

$$y'' = \frac{\mathrm{d}\left(\frac{\mathrm{d}y}{\mathrm{d}x}\right)}{\mathrm{d}x}.$$

Using the rule for finding the differential of a fraction, we get [50]:

$$y'' = \frac{\mathrm{d}^2y \cdot \mathrm{d}x - \mathrm{d}^2x \cdot \mathrm{d}y}{(\mathrm{d}x)^3}. \tag{13}$$

But by (9):

$$\mathrm{d}x = \varphi'(t)\, \mathrm{d}t, \qquad \mathrm{d}^2x = \varphi''(t)\, \mathrm{d}t^2$$

$$\mathrm{d}y = \psi'(t)\, \mathrm{d}t, \qquad \mathrm{d}^2y = \psi''(t)\, \mathrm{d}t^2.$$

Substituting in (13) and cancelling $\mathrm{d}t^3$, we finally have:

$$y'' = \frac{\psi''(t)\,\varphi'(t) - \varphi''(t)\,\psi'(t)}{[\varphi'(t)]^3}. \tag{14}$$

We remark that the expression (13) for y'' differs from the expression (3) of [55] for the same derivative ($n = 2$),

$$y'' = \frac{\mathrm{d}^2y}{\mathrm{d}x^2}, \tag{15}$$

this latter formula being obtained when x is the independent variable, whereas t is the independent variable in the parametric form of (9). When x is the independent variable, $\mathrm{d}x$ is treated as constant [50], i.e. independent of x, so that $\mathrm{d}^2x = \mathrm{d}(\mathrm{d}x) = 0$, being the differential of a constant. Formula (13) now reduces to (15).

Now that we can determine y' and y'', we can supply information regarding the direction of the tangent to the curve and its convexity and concavity, etc.

We take as an example the curve given by the equation:

$$x^3 + y^3 - 3axy = 0\,, \qquad (a > 0) \tag{16}$$

this being known as the "folium of Descartes".

We introduce a variable parameter t, putting:

$$y = tx\,, \tag{17}$$

and we consider the points of intersection of the curve with the straight line (17) that has variable slope t. Substituting for y from (17) in (16) and cancelling out x^2, we get:

$$x = \frac{3at}{1+t^3}\,,$$

whilst (17) now gives:

$$y = \frac{3at^2}{1+t^3}\,.$$

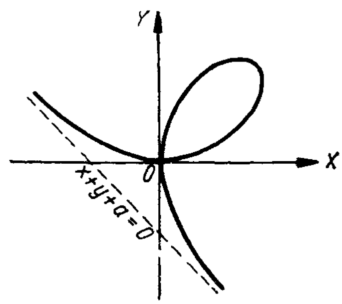

FIG. 85

These equations give the folium of Descartes in parametric form. We find the derivatives of x and y with respect to t:

$$\left.\begin{aligned} x_t' &= 3a\,\frac{(1+t^3) - 3t^2\cdot t}{(1+t^3)^2} = \frac{6a\left(\dfrac{1}{2} - t\right)}{(1+t^3)^2} \\ y_t' &= 3a\,\frac{2t(1+t^3) - 3t^2\cdot t^2}{(1+t^3)^2} = \frac{3at(2-t^3)}{(1+t^3)^2} \end{aligned}\right\} \tag{18}$$

We investigate the variation of x and y by dividing the total interval $(-\infty, +\infty)$ of variation of t into separate parts, within which x_t' and y_t' preserve invariable sign and do not tend to infinity. We thus note the values

$$t = -1,\ \ 0,\ \ \frac{1}{\sqrt[3]{2}}\ \text{and}\ \sqrt[3]{2}\,,$$

for which these derivatives vanish or become infinite. The signs of x_t' and y_t' within these intervals are found easily from (18); on calculating x and y at the ends of these intervals, we get the table below.

The curve corresponding to the table is shown in Fig. 85.

Interval of t	x'_t	y'_t	x	y
$(-\infty, -1)$	$+$	$-$	increases from 0 to ∞	decr. from 0 to $-\infty$
$(-1, 0)$	$+$	$-$	incr. from $-\infty$ to 0	decr. from $+\infty$ to 0
$(0, 1/\sqrt[3]{2})$	$+$	$+$	incr. from 0 to $\sqrt[3]{4a}$	incr. from 0 to $\sqrt[3]{2a}$
$(1/\sqrt[3]{2}, \sqrt[3]{2})$	$-$	$+$	decr. from $\sqrt[3]{4a}$ to $\sqrt[3]{2a}$	incr. from $\sqrt[3]{2a}$ to $\sqrt[3]{4a}$
$(\sqrt[3]{2}, +\infty)$	$-$	$-$	decr. from $\sqrt[3]{2a}$ to 0	decr. from $\sqrt[3]{4a}$ to 0

The slope of the tangent is given by the formula:

$$y' = \frac{y'_t}{x'_t} = \frac{t(2-t^3)}{2\left(\frac{1}{2} - t^3\right)} . \tag{19}$$

We notice that x and y are zero for $t = 0$ and $t = \infty$, and the curve cuts itself at the origin, as is clear from the figure.

Formula (19) gives us:

$$y'_x = 0 \quad \text{for} \quad t = 0 ,$$

$$y_x = \lim_{t\to\infty} \frac{t(2-t^3)}{2\left(\frac{1}{2} - t^3\right)} = \lim_{t\to\infty} \frac{t\left(\frac{2}{t^3} - 1\right)}{2\left(\frac{1}{2t^3} - 1\right)} = \infty \text{ for } t = \infty,$$

i.e. of the two branches of the curve, that intercept each other at the origin, one touches OX, and the other OY.

As t tends t),io ($-$ x and y tend to infinity, so that the curve has infinite branches. We find the asymptote:

$$\text{slope of asymptote} = \lim_{x\to\infty} \frac{y}{x} =$$

$$= \lim_{t\to-1} \frac{3at^2(1+t^3)}{3at(1+t^3)} = -1,$$

$$b = \lim_{t\to-1} (y+x) = \lim_{t\to-1} \frac{3at^2 + 3at}{1+t^3} = \lim_{t\to-1} \frac{6at+3a}{3t^2} = -a,$$

i.e. the equation of the asymptote is:

$$y = -x - a \quad \text{or} \quad x + y + a = 0 .$$

75. Van der Waal's equation. A gas that accurately satisfies Boyle's law and the Gay–Lussac law is known to satisfy the relation:

$$pv = RT ,$$

where T is absolute temperature, and R is a constant that is the same for all gases, provided one "gram-molecule" is taken, i.e. the number of grams of gas is equal to its molecular weight.

Gases do not strictly obey the above relationship in practice, and van der Waals gave a second expression that is a good deal more accurate. This formula reads:

$$\left(p + \frac{a}{v^2}\right)(v - b) = RT,$$

where a and b are positive constants that differ for different gases.

Solving the equation for p gives:

$$p = \frac{RT}{v - b} - \frac{a}{v^2}. \qquad (20)$$

We consider p as a function of v, with T constant, i.e. we take an isothermal change of state. We find the first derivative of p with respect to v:

$$\frac{\mathrm{d}p}{\mathrm{d}v} = -\frac{RT}{(v-b)^2} + \frac{2a}{v^3} =$$

$$\frac{1}{(v-b)^2}\left[\frac{2a(v-b)^2}{v^3} - RT\right]. \qquad (21)$$

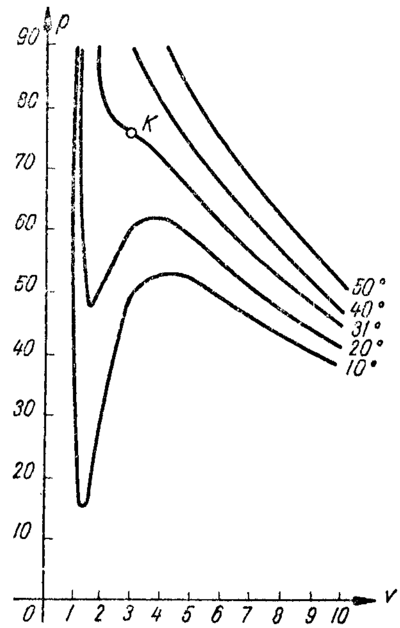

FIG. 86

We shall only take $v > b$. The reader is referred to a textbook of physics for the physical significance of this condition, as also of the curves that we shall obtain.

Equating the derivative to zero, we get:

$$\frac{2a\,(v-b)^2}{v^3} - RT = 0. \qquad (22)$$

We investigate the variation of the left-hand side of this equation as v varies from b to $(+\infty)$; thus we find its derivative with respect to v, recalling that RT is constant by hypothesis:

$$\left[\frac{2a\,(v-b)^2}{v^3}\right]' = 2a\,\frac{2(v-b)\,v^3 - 3v^2\,(v-b)^2}{v^6} = -\,\frac{2a(v-b)\,(v-3b)}{v^4},$$

whence it is clear that the derivative is positive for $b < v < 3b$, and negative for $v > 3b$, i.e. the left of (22) increases in the interval $(b, 3b)$, and decreases on further increase of v, whilst it reaches a maximum at $v = 3b$, equal to

$$\frac{8a}{27b} - RT.$$

It is easily seen by direct substitution that the left of (22) gives $(-RT)$ for $v = b$ and $v = +\infty$, and hence is negative. If the maximum obtained is also negative, i.e. if

$$RT > \frac{8a}{27b},$$

the left of (22) is always negative, and clearly, from (21), the derivative dp/dv will also always be negative, i.e. p decreases with increasing v.

On the other hand, if

$$RT < \frac{8a}{27b},$$

the left of (22) has a positive maximum for $v = 3b$, and (22) has one root v_1 in the interval $(b, 3b)$ and another root v_2 in $(3b, +\infty)$. The left of (22), and hence dp/dv, changes sign from minus to plus as v passes through v_1, i.e. a minimum p corresponds to this value of v. Similarly, $v = v_2$ gives a maximum p.

If, finally,

$$RT = \frac{8a}{27b}, \tag{23}$$

the maximum of the left of (22) is zero, and v_1, v_2 merge into the single value $v = 3b$; the left of (22), and dp/dv, preserve the minus sign on passage through this value, i.e. p is constantly decreasing with increasing v, so that $v = 3b$ gives the point of inflexion K of the curve. The values $v = v_k$, $p = p_k$, corresponding to the point of inflexion, along with $T = T_k$ as defined by (23), are called the critical volume, critical pressure, and critical temperature of the gas. The forms of the curve, corresponding to the three cases considered, are shown in Fig. 86.

76. Singular points of curves. We take the equation of a curve in implicit form:

$$F(x, y) = 0. \tag{24}$$

The slope of the tangent to the curve is given by [69]:

$$y' = -\frac{F'_x(x, y)}{F'_y(x, y)}, \tag{25}$$

where x, y are the coordinates of the point of contact.

We take the particular case of $F(x, y)$ being an integral polynomial in x and y. The curve (24) is called algebraic in this case. The partial derivatives $F'_x(x, y)$ and $F'_y(x, y)$ have fully determined values if the coordinates of a point M of curve (24) are substituted for x and y, and (25) defines the slope of the tangent in every case except those where the coordinates x, y, of the

point cause $F'_x(x, y)$ and $F'_y(x, y)$ to vanish. Any such point M is called a singular point of the curve (24).

A singular point of an algebraic curve (24) *is a point, the coordinates of which satisfy* (24) *and also the equations:*

$$F'_x(x, y) = 0 \,, \quad F'_y(x, y) = 0 \,. \tag{26}$$

In the case of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

condition (26) gives us $x = y = 0$; but (0,0) does not lie on the ellipse, which thus has no singular points. The same can be said of the hyperbola and parabola.

For the folium of Descartes:

$$x^3 + y^3 - 3axy = 0 \,,$$

condition (26) takes the form:

$$3x^2 - 3ay = 0 \quad \text{and} \quad 3y^2 - 3ax = 0 \,,$$

and it is immediately evident that the origin (0,0) is a singular point of the curve. We showed that the folium of Descartes cuts itself at the origin, the two intercepting branches of the curve having different tangents at this point: OX is the tangent to one branch, and OY the tangent to the other.

A singular point, where different branches of a curve intercept, each branch having its own particular tangent, is called a node.

For instance, the origin is a node of the folium of Descartes.

We give some further examples to show the various types of singular point of algebraic curves.

1. We take the curve:

$$y^2 - ax^3 = 0 \quad (a > 0) \,,$$

known as a *semicubical parabola.* It can easily be seen that the left-hand side of this equation, together with its partial derivatives with respect to x and y, vanishes at (0,0), so that the origin is a singular point of the curve. We draw the curve, so as to study it near this singular point. The explicit form of the equation is:

$$y = \pm \sqrt{ax^3} \,.$$

It is sufficient to draw the section of the curve corresponding to the plus sign, since the section with the minus sign is symmetrical as regards the first section about OX. It is clear from the equation that x cannot be less than zero, and that y increases from 0 to $(+\infty)$ as x increases from 0 to $(+\infty)$.

We find the first and second derivatives:

$$y' = \frac{3}{2}\sqrt{ax}; \qquad y'' = \frac{3\sqrt{a}}{4\sqrt{x}} \,.$$

We have $y' = 0$ for $x = 0$, and on noting that x can only tend to zero through positive values, we can say that OX is a tangent to the curve on

the right at the origin. It is also evident that y'' remains positive for the section of the curve in question in the interval $(0, +\infty)$, i.e. this section is concave on the side of positive ordinates.

Figure 87 shows the curve concerned, for $a = 1$. *Two branches of the curve approach the origin without going past it, the branches being on different sides of the same tangent at this singular point (and always on different sides in the present case)*. Such a singular point is called a *cusp of the first kind.*

2. We take the curve:

$$(y - x^2)^2 - x^5 = 0.$$

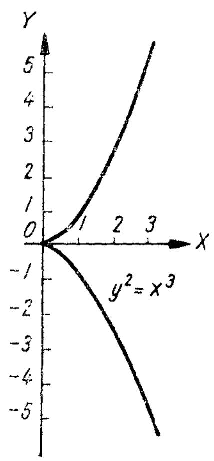

FIG. 87

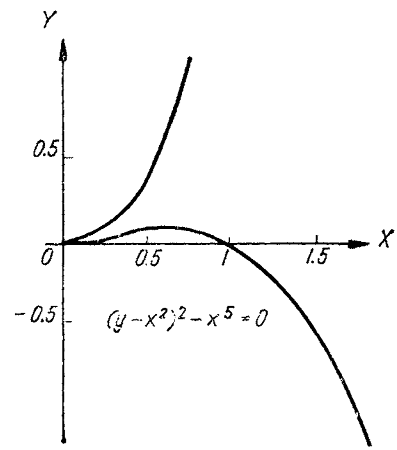

FIG. 88

It can easily be seen that the origin is a singular point. The explicit form of the equation is:

$$y = x^2 \pm \sqrt{x^5}.$$

It is clear from this equation that x can vary from 0 to $(+\infty)$. We find the first and second derivatives:

$$y' = 2x \pm \frac{5}{2}\sqrt{x^3}, \quad y'' = 2 \pm \frac{15}{4}\sqrt{x}$$

and investigate separately the two branches of the curve corresponding to the $(+)$ and $(-)$ signs.

We first of all remark that $y' = 0$ for $x = 0$ in both cases, so that, as in the previous example, OX is a tangent on the right for both branches.

We get the following results on investigating in the usual manner: y increases from 0 to $(+\infty)$ as x increases from 0 to $(+\infty)$ along the first branch, and the curve is concave; there is a peak (maximum) in the second branch at $x = 16/25$, a point of inflexion at $x = 64/225$, and a point of intersection with axis OX at $x = 1$.

All these data result in the curve shown in Fig. 88.

Two branches of the curve meet at the origin without going past it, have the same tangent at the origin, and are located on the same side of the tangent in the neighbourhood of this singular point. Such a singular point is called a *cusp of the second kind.*

3. We consider the curve:

$$y^2 - x^4 + x^6 = 0.$$

The origin is a singular point. The equation gives explicitly:

$$y = \pm x^2 \sqrt{1 - x^2}.$$

The implicit equation contains only even powers of x and y, so that the coordinate axes are axes of symmetry of the curve, and it is sufficient to take only the section of the curve corresponding to positive values of x and y. It is evident from the explicit equation that x can vary from (-1) to 1.

We find the first derivative:

$$y' = \frac{x(2 - 3x^2)}{\sqrt{1 - r^2}}.$$

We have $y = y' = 0$ for $x = 0$, i.e. the tangent at the origin coincides with OX; whilst for $x = 1$, $y = 0$ and $y' = \infty$, i.e. the tangent at (0,1) is parallel to OY. We find by the usual rule that the curve has a peak at $x = \sqrt{2/3}$.

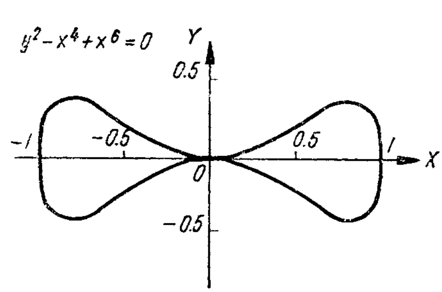

FIG. 89

The above data, including the symmetry, give us the curve of Fig. 89. *Two branches of the curve*, corresponding to the plus and minus signs before the square root, *touch each other* at the origin. A singular point of this sort is called a *point of osculation.*

4. We consider the curve:

$$y^2 - x^2(x - 1) = 0.$$

The origin is a singular point. The explicit form of the curve is:

$$y = \pm \sqrt{x^2 (x - 1)}.$$

Since the expression under the square root must not be negative, we can say that either $x = 0$ or $x \geqslant 1$.

We have $y = 0$ for $x = 0$. We now consider the branch corresponding to the plus sign. As x increases from 1 to $(+\infty)$, y increases from 0 to $(+\infty)$. It is evident from the expression for the first derivative:

$$y' = \frac{3x - 2}{2\sqrt{x - 1}},$$

that y' becomes infinite for $x = 1$, i.e. the tangent at (1,0) is parallel to OY. The second branch of the curve, corresponding to the minus sign, is

symmetrical with the first branch about OX. These data give us the curve shown in Fig. 90. In this case, whilst the coordinates of point O (0,0) satisfy the equation of the curve, *there are no other points of the curve in its vicinity. This type of singular point is called an isolated point.*

The above-mentioned types of singular point exhaust the possibilities as regards algebraic curves, though the coincidence of singular points of the same or different types is possible at certain points of algebraic curves.

Non-algebraic curves are referred to as *transcendental.*

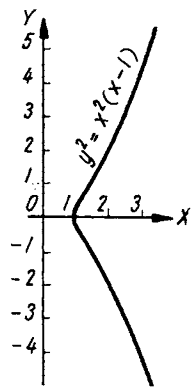

FIG. 90

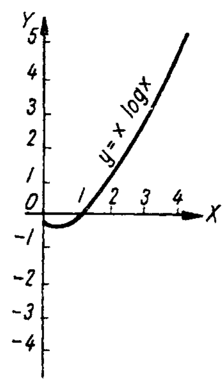

FIG. 91

We suggest that the reader might show that the curve of

$$y = x \log x$$

is as shown in Fig. 91. The origin is a *stop point* of the curve.

77. Elements of a curve. We give basic formulae connected with the concepts of tangent and curvature of a curve, and introduce some new concepts associated with the concept of tangent.

If the equation of the curve has the form:

$$y = f(x) , \tag{27}$$

the derivative $f'(x)$ of y with respect to x is the slope of the tangent, the equation of which can be written in the form:

$$Y - y = y'(X - x) \; ; \; (y' = f'(x)) , \tag{28}$$

where (x, y) are the coordinates of the point of contact, and (X, Y) are the current coordinates of the tangent. *The normal to a curve at any point of it (x, y) is the perpendicular drawn through the point to the tangent at the point.* We know from analytic geometry that the per-

pendicular to a line has a slope equal to the reciprocal with changed sign, i.e., the slope of the normal is $(-1/y')$, and the equation of the normal can be written as:

$$Y - y = -\frac{1}{y'}(X - x)$$

or

$$(X - x) + y'(Y - y) = 0. \tag{29}$$

Let M be any point of the curve, and let T and N be the points of intersection of the tangent and normal to the curve at M with axis OX; also, let Q be the base of the perpendicular dropped from M to OX (Fig. 92). *The segments QT and QN on OX are called respectively the subtangent and subnormal of the curve at M*; there are definite numbers corresponding to these segments, positive or negative, depending on their direction along OX. The lengths of segments MT and MN are referred to respectively as *the lengths of the tangent and normal to the curve at M*, these lengths always being reckoned positive. The abscissa of Q on OX is evidently equal to the abscissa x of M. Since T and N are the points of intersection of the tangent and normal with OX, their abscissae must be found by setting $Y = 0$ in the equations of the tangent and normal, then solving the equations obtained with respect to X. We thus get $(x - y/y')$ for the abscissa of T, and $(x + yy')$ for the abscissa of N. The magnitudes of the subtangent and subnormal are now easily found:

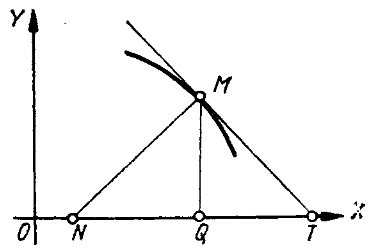

FIG. 92

$$\left.\begin{aligned} \overline{QT} &= \overline{OT} - \overline{OQ} = x - \frac{y}{y'} - x = -\frac{y}{y'}, \\ \overline{QN} &= \overline{ON} - \overline{OQ} = x + yy' - x = yy'. \end{aligned}\right\} \tag{30}$$

The lengths of the tangent and normal can now be found from the right-angled triangles MQT and MQN:

$$\begin{aligned} |\overline{MT}| &= \sqrt{\overline{MQ}^2 + \overline{QT}^2}\sqrt{y^2 + \frac{y^2}{y'^2}} = \pm\frac{y}{y'}\sqrt{1 + y'^2}, \\ |\overline{MN}| &= \sqrt{\overline{MQ}^2 + \overline{QN}^2}\sqrt{y^2 + y^2y'^2} = \pm y\sqrt{1 + y'^2}, \end{aligned} \tag{31}$$

where the sign $(\pm)$ must be chosen so that the expression on the right-hand side is positive.

We recall the formula for the radius of curvature of a curve [71]:

$$R = \pm \frac{(1 + y'^2)^{3/2}}{y''} . \tag{32}$$

Denoting the length of the normal by n, we get from the second of formulae (31):

$$\sqrt{1 + y'^2} = \pm \frac{n}{y},$$

and, on substituting this expression for $\sqrt{1 + y'^2}$ in (32), we get the following further expression for the radius of curvature:

$$R = \pm \frac{n^3}{y^3 y''} . \tag{32_1}$$

If the curve is given parametrically:

$$x = \varphi(t), \ y = \psi(t) ,$$

the first and second derivatives of y with respect to x are given by [74]:

$$y' = \frac{dy}{dx} = \frac{\psi'(t)}{\varphi'(t)},$$

$$y'' = \frac{d^2y dx - d^2x dy}{dy^3} = \frac{\psi''(t)\,\varphi'(t) - \varphi''(t)\,\psi'(t)}{[\varphi'(t)]^3} . \tag{33}$$

In particular, substituting this expression in (32), we get the following expression for the radius of curvature in this case:

$$R = \pm \frac{(dx^2 + dy^2)^{3/2}}{d^2y\, dx - d^2x\, dy} =$$

$$= \pm \frac{\{[\varphi'(t)]^2 + [\psi'(t)]^2\}^{3/2}}{\psi''(t)\,\varphi'(t - \varphi''(t)\,\psi'(t)} = \pm \frac{ds}{d\alpha}, \tag{34}$$

where α is the angle formed by the tangent with OX.

If the curve is given implicitly as

$$F(x, y) = 0 ,$$

we obtain the equation of the tangent from (25) as:

$$F'_x(x, y)\,(X - x) + F'_y(x, y)\,(Y - y) = 0. \tag{35}$$

78. The catenary. The curve whose equation is, with suitable choice of coordinates:

$$y = \frac{a}{2}\left(e^{x/a} + e^{-x/a}\right) \quad (a > 0)$$

is called the catenary. It gives the shape of a uniformly heavy string, hanging from its two ends. It is easily drawn by the rules of [73], and its form is shown in Fig. 93.

We find the first and second derivatives of y:

$$y' = \frac{1}{2}\left(e^{x/a} - e^{-x/a}\right),$$

$$y'' = \frac{1}{2a}\left(e^{x/a} + e^{-x/a}\right) = \frac{y}{a^2},$$

whence

$$1 + y'^2 = 1 + \frac{\left(e^{x/a} - e^{-x/a}\right)^2}{4} =$$

$$= \frac{4 + e^{2x/a} - 2 + e^{-2x/a}}{4} =$$

$$= \frac{\left(e^{x/a} + e^{-x/a}\right)^2}{4} = \frac{y^2}{a^2}.$$

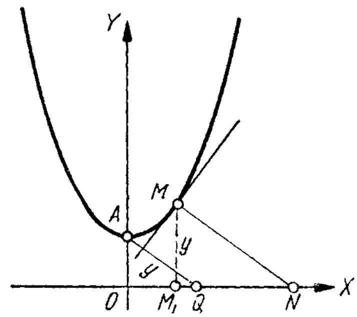

FIG. 93

Substituting this expression for $(1 + y'^2)$ in the second of formulae (31), we get for the length of the normal to the curve:

$$n = \frac{y^2}{a},$$

and substituting for n and y'' in (32_1), we get:

$$R = \frac{y^6 a^2}{a^3 y^3 y} = \frac{y^2}{a} = n,$$

i.e. *the radius of curvature of the catenary is equal to the length of the normal* MN. The ordinate has its least value $y = a$ at $x = 0$, and the corresponding point A of the curve is called its *vertex.*

Some further auxiliary lines are shown in the figure, which are needed by us later. The equation of the catenary is unchanged on substituting $(-x)$ for x, i.e. OY is an axis of symmetry of the curve.

79. The cycloid. We imagine a circle of radius a, rolling without slipping along a stationary straight line. The locus traced out by any point M of the circumference of the moving circle is called a *cycloid.*

We take the line on which the circle rolls as axis OX; we take as origin the initial position of the point M when this is the point of contact of the circle with OX, and we denote the angle of rotation of the circle by t. Further, we call the centre of the circle C, its point of contact with OX at a given position N; the base of the perpendicular dropped from M on to OX is called

Q, and the base of the perpendicular dropped from M on to the diameter NN_1 of the circle is called R (Fig. 94).

Noting that, with no slipping:

$$\overline{ON} = \text{arc } NM = at ,$$

we can write the coordinates of M, describing the cycloid, in terms of parameter t = angle NCM:

$$x = \overline{OQ} = \overline{ON} - \overline{QN} = at - a \sin t = a(t - \sin t) ,$$

$$y = \overline{QM} = \overline{NC} - \overline{RC} = a - a \cos t = a(1 - \cos t).$$

This gives the cycloid in parametric form.

We remark first of all that it is sufficient to consider t varying in the interval $(0, 2\pi)$, which corresponds to a full turn of the circle. After this

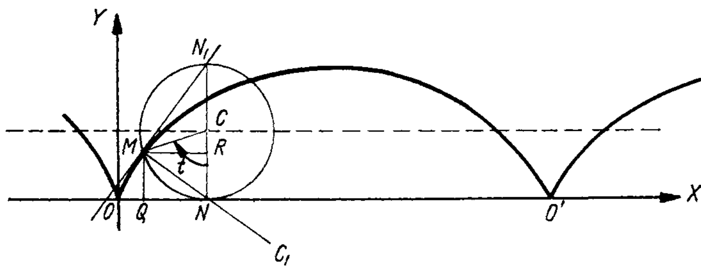

FIG. 94

full turn, M again coincides with the point of contact O' of the circle with OX, having simply shifted along $\overline{OO'} = 2\pi a$. The section of the curve obtained with a further revolution will be exactly like arc OO', being obtained by moving this arc along by an amount $2\pi a$ to the right, and so on. We now find the first and second derivatives of x and y with respect to t:

$$\left.\begin{aligned} &\frac{dx}{dt} = \varphi'(t) = a(1 - \cos t), \quad &&\frac{dy}{dt} = \psi'(t) = a \sin t, \\ &\frac{d^2x}{dt^2} = \varphi''(t) = a \sin t, &&\frac{d^2 y}{dt^2} = \psi''(t) = a \cos t. \end{aligned}\right\} \tag{36}$$

By the first of formulae (33), the slope of the tangent is:

$$y' = \frac{a \sin t}{a(1 - \cos t)} = \frac{2 \sin \frac{1}{2} t \cos \frac{1}{2} t}{2 \sin^2 \frac{1}{2} t} = \cot \frac{1}{2} t .$$

This formula leads to a simple method of constructing the tangent to a cycloid. We associate the point N_1 with point M of the circle. Angle MN_1N is the angle subtended at the circumference by arc $NM = t$, and is therefore equal to $t/2$. We have from the right-angled triangle RMN_1 (Fig. 94):

$$\angle RMN_1 = \frac{1}{2}\pi - \frac{1}{2}t, \qquad \tan \angle RMN_1 = \cot \frac{1}{2}t.$$

On comparing this with the expression for y', we see that MN_1 is a tangent to the cycloid, i.e.:

In order to construct the tangent at any point M of a cycloid, it is sufficient to join this point to the end N_1 of the diameter whose other end is the point of contact of the rolling circle with the axis OX.

The line MN, joining M to the other end of this diameter of the circle, is perpendicular to MN_1, since angle N_1MN is subtended by a diameter;

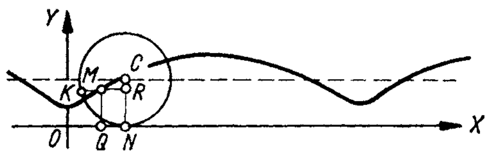

FIG. 95

hence *MN is the normal to the cycloid.* The length of the normal, $n = \overline{MN}$, is found directly from the right-angled triangle N_1MN:

$$n = 2a \sin \frac{1}{2} t.$$

We obtain the radius of curvature of the cycloid from (34) and (36):

$$R = \pm \frac{[a^2(1-\cos t)^2 + a^2 \sin^2 t]^{3/2}}{a \cos t \times a(1-\cos t) - a \sin t \times a \sin t} =$$

$$= \pm \frac{a(2 - 2\cos t)^{3/2}}{\cos t - 1} = 2^{3/2} \times a(1-\cos t)^{1/2} = 4a \sin \frac{1}{2} t.$$

We leave only the plus sign standing in the last expression, since t lies in $(0, 2\pi)$ for the first branch of the cycloid, and $\sin t/2$ cannot be negative.

Comparing this expression with that for the length of the normal n, we have $R = 2n$, i.e. *the radius of curvature of a cycloid is equal to twice the length of the normal* ($\overline{MC}_1$ in Fig. 94).

If the point M that described the cycloid were to lie inside, or outside, instead of on, the circle, the corresponding curve obtained when the circle rolled would be a *curtate* or *prolate cycloid* (both curves are sometimes called *trochoids*).

Let h denote the distance of M from the centre C of the rolling circle. The rest of the notation is the same. We first take $h < a$, i.e. the case of M lying inside the circle (Fig. 95). We have directly from the figure:

$$x = \overline{OQ} = \overline{ON} - \overline{QN} = at - h \sin t; \; y = \overline{QM} = \overline{NC} - \overline{RC} = a - h \cos t.$$

The equations are the same for $h < a$, but the curve has the form shown in Fig. 96.

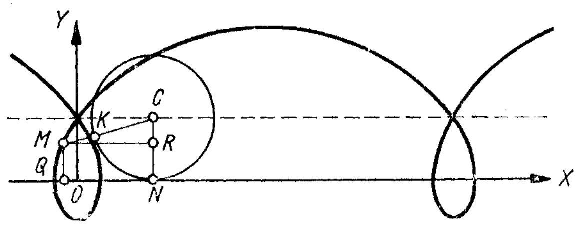

FIG. 96

80. Epicycloid and hypocycloid. If the circle carrying the point M rolls without slipping on another circle, instead of on a straight line, two general classes of curves are obtained: *epicycloids*, when the rolling circle is outside, and *hypocycloids*, when it is inside, the fixed circle.

We find the equation of the *epicycloid*. We take the centre of the fixed circle as origin; axis OX is taken along the line joining this centre O to K, the initial position of point M, where the two circles originally touch each

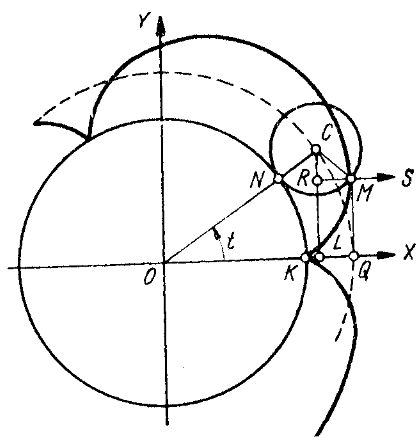

FIG. 97

other. Let a be the radius of the rolling circle, and b the radius of the fixed circle; let parameter t be the angle between OX and radius ON of the fixed circle, where N is the point of contact of the rolling circle after it has turned through an angle $\varphi = \angle NCM$ (Fig. 97).

Since the circle rolls without slipping, we can write:

$$\text{arc } KN = \text{arc } NM ,$$

i.e.

$$bt = a\varphi , \quad \varphi = \frac{bt}{a} .$$

We find directly from the figure:

$$\left.\begin{aligned} x &= \overline{OQ} = \overline{OL} + \overline{LQ} = \overline{OC}\cos\angle KOC - \overline{CM}\cos\angle SMC = \\ &= (a+b)\cos t - a\cos(t+\varphi) = (a+b)\cos t - a\cos\frac{a+b}{a}t, \\ y &= \overline{QM} = \overline{LC} - \overline{RC} = \overline{OC}\sin\angle KOC - \overline{CM}\sin\angle SMC = \\ &= (a+b)\sin t - a\sin(t+\varphi) = (a+b)\sin t - a\sin\frac{a+b}{a}t. \end{aligned}\right\} \quad (37)$$

The curve consists of a series of identical arcs, each corresponding to a complete revolution of the moving circle, i.e. to an increase of φ by 2π, and of t by $2a\pi/b$

The ends of the arcs thus correspond to:

$$t = 0 , \frac{2a\pi}{b} , \frac{4a\pi}{b} , \ldots , \frac{2pa\pi}{b} , \ldots$$

A necessary and sufficient condition that we should eventually arrive at the initial point K of the curve is that one of these ends should coincide with K, i.e. that there should exist integers p and q satisfying

$$\frac{2pa\pi}{b} = 2q\pi ,$$

since a certain number of complete turns about O corresponds to K. The above condition can be written:

$$\frac{a}{b} = \frac{q}{p} .$$

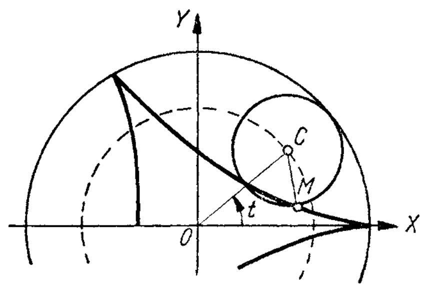

FIG. 98

Such numbers p and q exist if, and only if, a and b are commensurable with each other; otherwise, a/b is irrational and cannot be equal to the ratio of two integers.

It thus follows that the epicycloid represents a closed curve if, and only if, the radius of the moving circle is commensurable with that of the fixed circle; otherwise, the curve is not closed, and it is impossible ever to return to the starting-point K.

This remark also applies to the hypocycloid (Fig. 98), the equation of which can be obtained by simply replacing a by $(-a)$ in the equation of the epicycloid:

$$\left.\begin{aligned} x &= (b-a)\cos t + a\cos\frac{b-a}{a}t, \\ y &= (b-a)\sin t - a\sin\frac{b-a}{a}t. \end{aligned}\right\} \quad (38)$$

We note some particular cases. Let $b = a$ in the case of the epicycloid, i.e. the radii of the fixed and moving circles are equal. We get a curve consisting of a single branch in this case (Fig. 99), its equation being, by substitution of $b = a$ in (37):

$$x = 2a\cos t - a\cos 2t,$$

$$y = 2a\sin t - a\sin 2t.$$

This curve is called a *cardioid*.

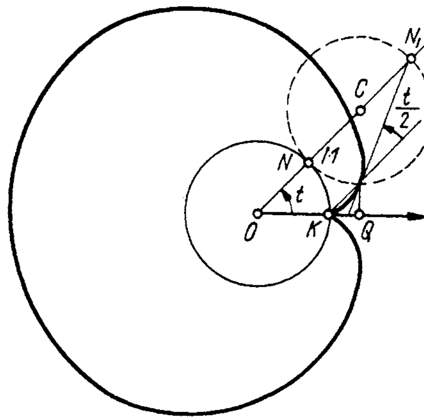

FIG. 99

We find the distance r of point $M(x, y)$ of this curve from the point K, with coordinates $(a, 0)$; for this purpose, we write expressions for $(x - a)$ and y in a more convenient form:

$$x - a = 2a\cos t - a(\cos^2 t - \sin^2 t) - a = 2a\cos t - 2a\cos^2 t =$$

$$= 2a\cos t(1 - \cos t),$$

$$y = 2a\sin t - 2a\sin t\cos t = 2a\sin t(1 - \cos t),$$

whence

$$r = |\overline{KM}| = \sqrt{(x-a)^2 + y^2} =$$

$$= \sqrt{4a^2\cos^2 t(1 - \cos t)^2 + 4a^2\sin^2 t(1 - \cos t)^2} = 2a(1 - \cos t).$$

Clearly, $(x - a)$ and y are the projections of KM on the x and y axes, whilst they are also equal, as is seen from the expressions above, to the length $\overline{KM}$ multiplied by $\cos t$ and $\sin t$ respectively, and hence it follows that KM forms an angle t with the positive direction of OX, i.e. is parallel to the radius ON. This result becomes important later, when giving a rule for constructing the tangent to a cardioid.

We introduce angle $\theta = \pi - t$, formed by KM with the negative direction of OX. We now have for r:

$$r = 2a(1 + \cos\theta).$$

This is the polar equation of the cardioid; the curve will be considered in more detail when polar coordinates are discussed.

We now note some particular cases of hypocycloids. Setting $b = 2a$ in (38), we get:

$$x = 2a\cos t = b\cos t\,, \quad y = 0\,,$$

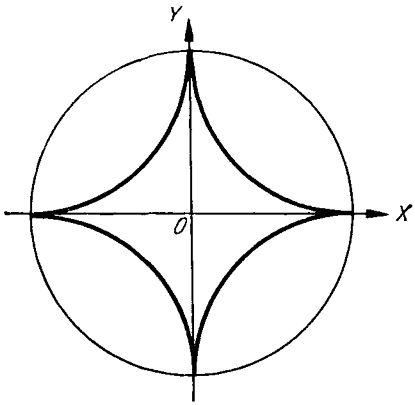

FIG. 100

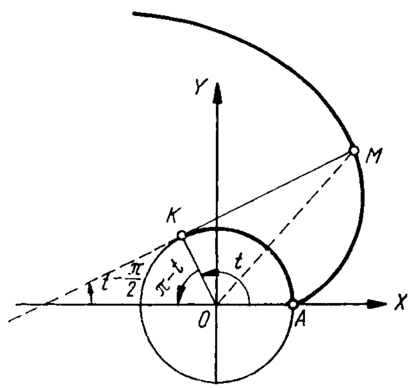

FIG. 101

i.e. if the radius of the fixed circle is twice the radius of the rolling circle, the point M moves along a diameter of the fixed circle.

We now take $b = 4a$. The hypocycloid in this case consists of four branches (Fig. 100), being called an *astroid* in this particular case. With $b = 4a$, (38) now gives:

$$x = 3a\cos t + a\cos 3t = 3a\cos t + a(4\cos^3 t - 3\cos t) =$$

$$= 4a\cos^3 t = b\cos^3 t\,,$$

$$y = 3a\sin t - a\sin 3t = 3a\sin t - a(3\sin t - 4\sin^3 t) =$$

$$= 4a\sin^3 t = b\sin^3 t.$$

We can eliminate the parameter t by taking the 2/3 power of each side of the above equations, then adding the equations term by term; this gives us the implicit equation of the astroid:

$$x^{2/3} + y^{2/3} = b^{2/3}.$$

81. Involute of a circle. This is the name of the curve described by the end M of a flexible string, gradually unwound from a fixed circle of radius a, so that it remains tangential to the circle at the point K where it leaves the circle (Fig. 101).

Taking as the parameter t the angle between the positive direction of OX and the radius drawn to the point K, and noting that $\overline{KM}$ = arc $AK = at$, we obtain the equation of the involute of a circle in parametric form:

$$x = \text{proj}_{OX}\,\overline{OM} = \text{proj}_{OX}\,\overline{OK} + \text{proj}_{OX}\,\overline{KM} = a\cos t + at\sin t\,,$$

$$y = \text{proj}_{OY}\,\overline{OM} = \text{proj}_{OY}\,\overline{OK} + \text{proj}_{OY}\,\overline{KM} = a\sin t - at\cos t\,.$$

We use the first of formulae (33) to find the slope of the tangent:

$$y' = \frac{a\cos t - a\cos t + at\sin t}{-a\sin t + a\sin t + at\cos t} = \tan t.$$

The slope of the normal to the involute of a circle is thus:

$$-\cot t = \tan\left(t - \frac{1}{2}\pi\right),$$

whence it is clear that MK is the normal to the involute. This property holds good, as we shall see later, for the involute of any curve.

82. Curves in polar coordinates. The position of a point M on the plane (Fig. 102) is defined in polar coordinates: (1) by its distance r from a given point O (the pole), and (2) by the angle θ between the direction of OM and some given direction L (the polar axis). It is usual to refer to r as the *radius vector* and to θ as the *polar angle*. If the polar axis is taken as axis OX, and the pole O as origin, we obviously have (Fig. 103):

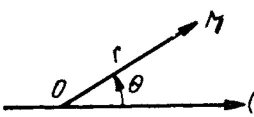

FIG. 102

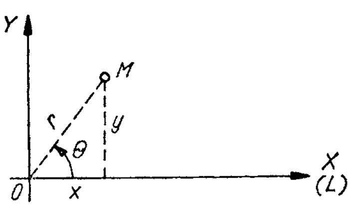

FIG. 103

$$x = r\cos\theta,\ y = r\sin\theta. \qquad (39)$$

To a given position of the point M there corresponds a single determinate positive value of r, but an infinite number of values of θ, differing by multiples of 2π. If M coincides with O, $r = 0$, and θ is completely indeterminate.

Any functional relationship of the form $r = f(\theta)$ (explicit) or $F(r, \theta) = 0$ (implicit) has a corresponding graph in the polar system of coordinates. The explicit form is more commonly encountered:

$$r = f(\theta). \qquad (40)$$

We shall later consider negative, as well as positive, values of r, it being agreed to take r in the opposite direction to that corresponding to θ, in the case when the value of r corresponding to θ is negative.

Assuming that r is a function of θ for a given curve, equations (39) are seen to represent the parametric equations of this curve, where x and y depend on the parameter θ both directly and through the medium of r. We can thus apply formulae (33) and (34) [77] in this case. On letting a denote the angle formed by the tangent with axis OX, we have by using the first of formulae (33):

$$\tan a = y' = \frac{r' \sin \theta + r \cos \theta}{r' \cos \theta - r \sin \theta},$$

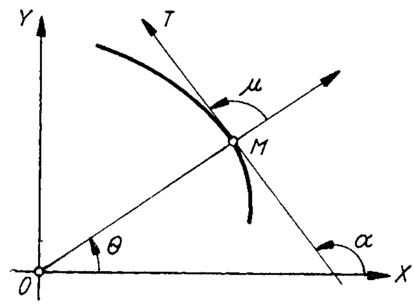

FIG. 104

where r' denotes the derivative of r with respect to θ.

We now further introduce μ, the angle between the positive direction of the radius vector and the tangent to the curve (Fig. 104). We have:

$$\mu = a - \theta,$$

and hence:

$$\cos \mu = \cos a \cos \theta + \sin a \sin \theta,$$

$$\sin \mu = \sin a \cos \theta - \cos a \sin \theta.$$

Differentiating equations (39) with respect to s, and noting that $\mathrm{d}x/\mathrm{d}s$ and $\mathrm{d}y/\mathrm{d}s$ are respectively equal to $\cos a$ and $\sin a$, we get:

$$\cos a = \cos \theta \frac{\mathrm{d}r}{\mathrm{d}s} - r \sin \theta \frac{\mathrm{d}\theta}{\mathrm{d}s}, \quad \sin a = \sin \theta \frac{\mathrm{d}r}{\mathrm{d}s} + r \cos \theta \frac{\mathrm{d}\theta}{\mathrm{d}s}.$$

Substituting these expressions for $\cos a$ and $\sin a$ in the above expressions for $\cos \mu$ and $\sin \mu$, we have:

$$\cos \mu = \frac{\mathrm{d}r}{\mathrm{d}s}, \quad \sin \mu = \frac{r\,\mathrm{d}\theta}{\mathrm{d}s} \tag{41}$$

and hence:

$$\tan \mu = \frac{r\,\mathrm{d}\theta}{\mathrm{d}r} = \frac{r}{\mathrm{d}r/\mathrm{d}\theta} = \frac{r}{r'}. \tag{41$_1$}$$

It follows from (39):

$$\mathrm{d}x = \cos \theta \,\mathrm{d}r - r \sin \theta \,\mathrm{d}\theta,$$

$$\mathrm{d}y = \sin \theta \,\mathrm{d}r + r \cos \theta \,\mathrm{d}\theta,$$

and thus:

$$ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{(dr)^2 + r^2 (d\theta)^2}\,; \tag{42}$$

using $\alpha = \mu + \theta$, then dividing numerator and denominator by $d\theta$, we also have:

$$R = \pm \frac{ds}{d\alpha} = \frac{[(dr)^2 + r^2 (d\theta)^2]^{1/2}}{d\mu + d\theta} = \pm \frac{(r^2 + r'^2)^{1/2}}{1 + \dfrac{d\mu}{d\theta}}.$$

We get by using (41_1):

$$\mu = \text{arc tan}\,\frac{r}{r'}, \quad \frac{d\mu}{d\theta} = \frac{1}{1 + \left(\dfrac{r}{r'}\right)^2} \cdot \frac{r'^2 - rr''}{r'^2} = \frac{r'^2 - rr''}{r^2 + r'^2},$$

where r' and r'' are the first and second derivatives of r with respect to θ. Substituting this expression for $d\mu/d\theta$ in the above expression for R, we have:

$$R = \pm \frac{(r^2 + r'^2)^{3/2}}{r^2 + 2r'^2 - rr''}. \tag{43}$$

83. Spirals. We can distinguish three types of spiral:

the spiral of Archimedes: $r = a\theta$,
the hyperbolic spiral: $r\theta = a$, $(a > 0; b > 0)$,
the logarithmic spiral: $r = be^{a\theta}$.

The spiral of Archimedes has the form indicated in Fig. 105, with the dotted curve corresponding to $\theta < 0$. Negative values of r correspond to negative values of θ, i.e. a negative r must be taken in the opposite direction to that defined by the θ in question.

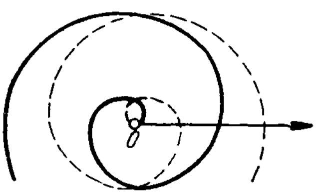

FIG. 105

Every radius vector cuts the curve an infinite number of times, the distance between any two successive points of intersection being constant and equal to $2a\pi$. This is evident from the fact that the direction of the radius vector corresponding to a given value of θ does not change if θ is increased by 2π, 4π, ...; whilst the length of r, given by $r = a\theta$, receives increments of $2a\pi$, $4a\pi$, ...

The *hyperbolic spiral* is shown in Fig. 106. We assume $\theta > 0$, and consider what happens to the curve as θ tends to 0. It is clear from the equation

$$r = \frac{a}{\theta}$$

that r now tends to infinity. We take some point M of the curve with sufficiently small θ, and let MQ be the perpendicular from M to the polar axis.

We get from the right-angled triangle MOQ (Fig. 106):

$$\overline{QM} = r \sin\theta = \frac{a \sin\theta}{\theta},$$

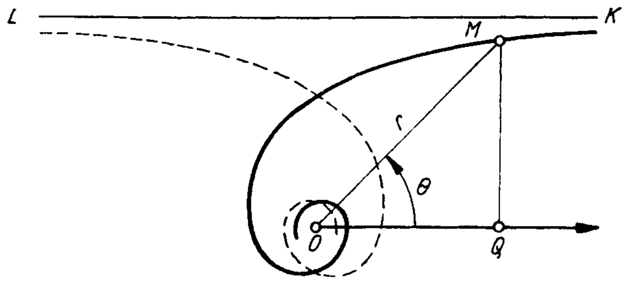

FIG. 106

and as θ tends to zero:

$$\lim_{\theta\to 0} \overline{QM} = \lim_{\theta\to 0} a \frac{\sin\theta}{\theta} = a.$$

The distance between the point M of the curve and the polar axis thus tends to a as θ tends to zero, so that the curve has an asymptote LK, parallel to the polar axis and at a distance a from it.

We see further that r does not vanish for any finite θ, but only tends to zero as θ tends to infinity. The curve thus indefinitely approaches the pole O whilst spiralling round it, and unlike the spiral of Archimedes, never actually reaches O. Such a point is in general referred to as an *asymptotic point* of a curve.

The *logarithmic spiral* is shown in Fig. 107.

We have $r = b$ for $\theta = 0$, whilst r tends to $(+\infty)$ for $\theta \to (+\infty)$; further, r tends to zero without vanishing as $\theta \to (-\infty)$. We have in this case:

$$r' = abe^{a\theta} \quad \text{and} \tan\mu = \frac{r}{r'} = \frac{1}{a},$$

i.e. *the angle μ between the tangent and radius vector is constant at any point of the logarithmic spiral.*

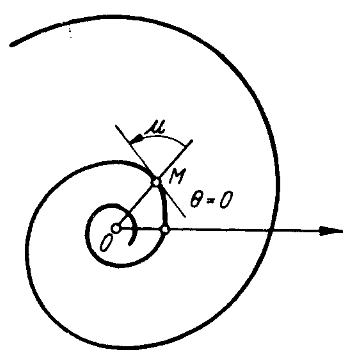

FIG. 107

84. The limaçon and cardioid. We draw a circle of diameter $OA = 2a$ (Fig. 108); we draw a radius vector through vector point O of the circumference to cut the circle in a point D, and produce it further to M, such that the distance $\overline{DM} = h$ is constant. The locus of M is in general called a limaçon.

Noting that:

$$\overline{OD} = 2a \cos\theta \quad \text{and} \quad \overline{OM} = r,$$

the equation of the limaçon will be:

$$r = 2a \cos \theta + h.$$

If $h > 2a$, this equation gives only positive values of r, and the curve is as shown in Fig. 109. If $h < 2a$, r also takes negative values, and the curve

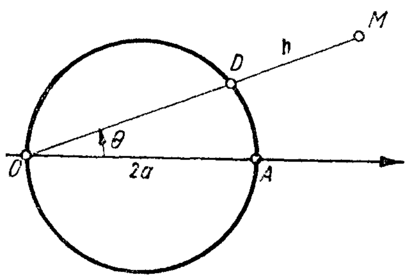

FIG. 108

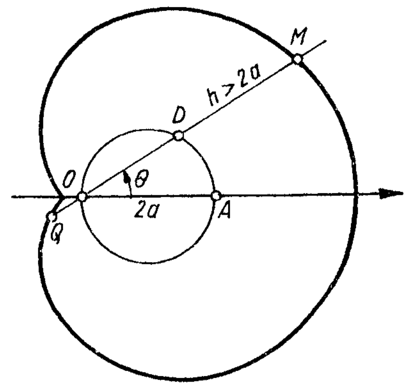

FIG. 109

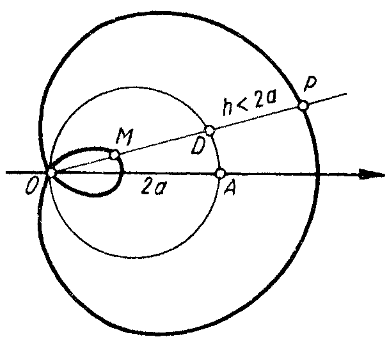

FIG. 110

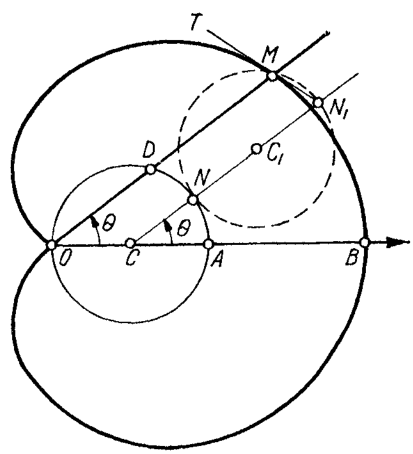

FIG. 111

has the form shown in Fig. 110. The curve cuts itself at the point O. Finally, for $h = 2a$, the equation of the limaçon becomes:

$$r = 2a(1 + \cos \theta) ,$$

i.e. the limaçon now becomes a cardioid [80], differing from that of [80] merely in its disposition (Fig. 111). We have $r = 0$ for $\theta = \pi$, i.e. the curve passes through O.

We find the first and second derivatives of r with respect to θ.

$$r' = -2a \sin \theta \,, \quad r'' = -2a \cos \theta \,.$$

We calculate $\tan \mu$:

$$\tan \mu = \frac{r}{r'} = \frac{2a\,(1 + \cos \theta)}{-2a \sin \theta} = -\cot \frac{1}{2}\theta = \tan \left(\frac{1}{2}\pi + \frac{1}{2}\theta\right),$$

i.e.

$$\mu = \frac{1}{2}\pi + \frac{1}{2}\theta \,. \tag{44}$$

As was shown previously [80], the cardioid can be obtained as the curve described by a point of a circle, rolling on the circle mentioned above of diameter $OA = 2a$, the diameter of the rolling circle being equal to that of the fixed circle. Let C be the centre of the fixed circle, M be any point of the cardioid, N be the point of contact of the rolling circle at the position corresponding to M, and NN_1 be the diameter of the moving circle (Fig. 111). We saw above [80] that the lines OM and CN_1 are parallel† i.e. angle $ACN = \theta$, and hence:

$$\text{arc } NM = \text{arc } ON = \pi - \theta \,.$$

Angle MN_1N subtended at the circumference by arc NM is equal to $\pi/2 - \theta/2$, and finally the angle between the directions OM and N_1M is equal to:

$$\pi - \left(\frac{1}{2}\pi - \frac{1}{2}\theta\right) = \frac{1}{2}\pi + \frac{1}{2}\theta = \mu \,,$$

whence it is clear that N_1M is the tangent to the cardioid at M. We thus get the following rule:

To draw the tangent to a cardioid at any point M of it, it is sufficient to join this point to the end N_1 of the diameter of the rolling circle, the other end of which is the point of contact of the rolling with the fixed circle ; the normal is then along the line MN.

The rule given above for constructing the tangent to a cardioid is obtained very simply from kinematic considerations. It is known that, in general, the motion of a constrained system in a plane occurs at any given moment by rotation about a fixed point (the instantaneous centre), the position of this point in general changing with the course of time. In the case of the rolling shown in Fig. 111, the instantaneous centre is the point of osculation N of the rolling with the fixed circle, so that the velocity of the point M, in the direction of the tangent to the cardioid, must be perpendicular to NM, i.e. NM must be normal to the cardioid, whilst the perpendicular to it, N_1M, must be tangential to the cardioid. It follows from these remarks that the rule given for constructing the tangent must be generally applicable to the curves described by any given point of a circle which rolls without slipping on a fixed circle.

† These two lines were KM and ON_1 in [80] (Fig. 99).

85. Cassini's ovals and the lemniscate. An oval of Cassini is the locus traced out by a point M, moving so that the product of its distances from two given points F_1 and F_2 is constant:

$$\overline{F_1M} \cdot \overline{F_2M} = b^2.$$

Let the length of $\overline{F_1F_2}$ be $2a$, let $\overline{F_1F_2}$ be directed along the polar axis, and let the pole O be at the mid-point of $\overline{F_1F_2}$.

We find from triangles OMF_1 and OMF_2 (Fig. 112):

$$\overline{F_1M^2} = r^2 + a^2 + 2ar\cos\theta\,,$$

$$\overline{F_2M^2} = r^2 + a^2 - 2ar\cos\theta\,.$$

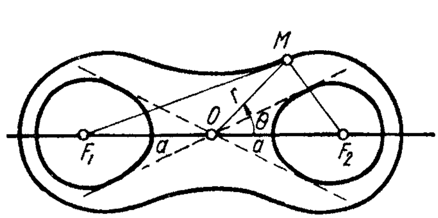

FIG. 112

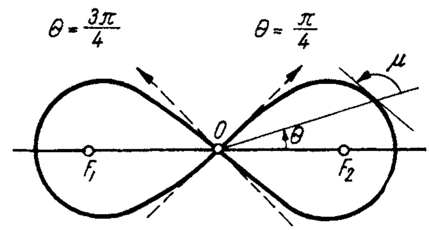

FIG. 113

Substituting these expressions in the equation for an oval, then squaring both sides, we get after simple re-arrangement:

$$r^4 - 2a^2 r^2 \cos 2\theta + a^4 - b^4 = 0\,,$$

whence

$$r^2 = a^2 \cos 2\theta \pm \sqrt{a^4 \cos^2 2\theta - (a^4 - b^4)}\,.$$

The cases corresponding to $a^2 < b^2$ and $a^2 > b^2$ are shown in Fig. 112, the second case giving the curve consisting of two separate closed loops. We give a detailed discussion only of the important case when $a^2 = b^2$. The corresponding curve is called a *lemniscate*, its equation being:

$$r^2 = 2a^2 \cos 2\theta\,.$$

This equation only gives real values of r when $\cos 2\theta \geqslant 0$, i.e. when θ lies in one of the intervals:

$$\left(0\,, \frac{\pi}{4}\right), \quad \left(\frac{3\pi}{4}\,, \frac{5\pi}{4}\right), \quad \left(\frac{7\pi}{4}\,, 2\pi\right),$$

whilst r vanishes for

$$\theta = \frac{\pi}{4}\,,\ \frac{3\pi}{4}\,,\ \frac{5\pi}{4}\,,\ \frac{7\pi}{4}\,.$$

The curve is easily constructed on the basis of these data (Fig. 113).

The curve cuts itself at O, and the dotted lines show the tangents to the two branches of the curve intercepting at O. We get by differentiating both sides of the equation of the lemniscate with respect to θ:

$$2rr' = -4a^2 \sin 2\theta \quad \text{or} \quad r' = -\frac{2a^2 \sin 2\theta}{r},$$

whence

$$\tan \mu = \frac{r}{r'} = -\frac{r^2}{2a^2 \sin 2\theta} = -\frac{2a^2 \cos 2\theta}{2a^2 \sin 2\theta} = -\cot 2\theta = \tan\left(\frac{1}{2}\pi + 2\theta\right),$$

$$\mu = \frac{1}{2}\pi + 2\theta.$$

To pass from polar to rectangular coordinates, we have from (39):

$$r^2 = x^2 + y^2, \ \cos \theta = \frac{x}{r}, \ \sin \theta = \frac{y}{r}.$$

On writing the equation of the lemniscate in the form:

$$r^2 = 2a^2(\cos^2 \theta - \sin^2 \theta)$$

and substituting the above expressions, we obtain the equation of the lemniscate in rectangular coordinates:

$$x^2 + y^2 = 2a^2 \frac{x^2 - y^2}{x^2 + y^2} \quad \text{or} \quad (x^2 + y^2)^2 = 2a^2 (x^2 - y^2),$$

whence it is clear that the lemniscate is a fourth order algebraic curve.

EXERCISES ON CHAPTER II

1. Find the increment in the function $y = x^2$ corresponding to the following changes in the argument: (*a*) from $x = 1$ to $x_1 = 2$; (*b*) from $x = 1$ to $x_1 = 1.1$; (*c*) from $x = 1$ to $x = 1 + h$.
2. Find Δy for the function $y = \sqrt[3]{x}$ if (*a*) $x = 0$, $\Delta x = 0.001$; (*b*) $x = 8$, $\Delta x = -9$; (*c*) $x = a$, $\Delta x = h$.
3. Why is it that for the function $y = 2x + 3$ we can find Δy corresponding to an increment $\Delta x = 5$ in x, but we cannot find Δy for the function $y = x^2$?
4. Find the increment Δy and the ratio $\Delta y/\Delta x$ for the functions: (*a*) $y = (x^2 - 2)^{-2}$ at $x = 1$ with $\Delta x = 0.4$; (*b*) $y = \sqrt{x}$ at $x = 0$ with $\Delta x = -0.0001$; (*c*) $y = \log_{10} x$ at $x = 100{,}000$ with $\Delta x = -90{,}000$.
5. Find Δy and $\Delta y/\Delta x$ corresponding to a change in the argument from x to $x + \Delta x$: (*a*) $y = ax + b$; (*b*) $y = x^3$; (*c*) $y = x^{-2}$; (*d*) $y = \sqrt{x}$; (*e*) $y = 2^x$; (*f*) $y = \log x$.
6. Find the slope of the chord of the parabola $y = 2x - x^2$, the abscissae of the end-points of which are (*a*) $x_1 = 1$, $x_2 = 2$; (*b*) $x_1 = 1, x_2 = 0.9$; (*c*) $x_1 = 1$, $x_2 = 1 + h$. To what limit does the slope of the chord tend in the last case as $h \to 0$.

7. The motion of a point is governed by the law $s = 2t^2 + 3t + 5$, where the distance s is measured in centimetres and the time t in seconds. What is the value of the average speed of the point in the time interval from $t = 1$ to $t = 5$?
8. Find the average rise in the curve $y = 2^x$ in the interval $1 \leqslant x \leqslant 5$.
9. Find the average rise in the curve $y = f(x)$ in the interval $(x, x + \Delta x)$.
10. What is understood by the statement that the slope of the curve $y = f(x)$ is given at the point x?
11. Give a definition of (*a*) average angular velocity, (*b*) instantaneous angular velocity.
12. In a heated body whose temperature is changing with time how should we define (*a*) average rate of change of temperature (*b*) instantaneous rate of change of temperature?
13. What should we understand by rate of change of reacting matter in a chemical reaction?
14. Let $m = f(x)$ be the non-homogeneous distribution of mass in a bar occupying the interval $(0, x)$. What should be understood by (*a*) the mean linear density of matter in the interval $(x, x + \Delta x)$; (*b*) the linear density of matter at the point x.
15. Find the ratio $\Delta y/\Delta x$ for the function $y = 1/x$ at the point $x = 2$ if (*a*) $\Delta x = 1$; (*b*) $\Delta x = 0.1$; (*c*) $\Delta x = 0.01$. What is the value of the derivative $y'(2)$?
16. Find the derivative of $y = \tan x$.
17. Find the derivatives of the functions (*a*) $y = x^3$; (*b*) $y = 1/x^2$; (*c*) $y = \sqrt{x}$, (*d*) $y = \cot x$.
18. Calculate $f'(8)$ if $f(x) = \sqrt[3]{x}$.
19. Find $f'(0), f'(1), f'(2)$ if $f(x) = x(x - 1)^2 (x - 2)^3$.
20. The motion of a point is governed by $s = 5t^2$ where s is given in metres and t in seconds. Find the velocity of the point when $t = 3$.
21. What are the values of the slopes of the curves $y = 1/x$ and $y = x^2$ at the point at which they intersect? Find the angle between the tangents to the curves at the point of intersection.
22. State which of the following functions do not possess finite derivatives at the points stated: (*a*) $y = \sqrt[3]{x^2}$ at $x = 0$, (*b*) $y = \sqrt[3]{x - 1}$ at $x = 1$; (*c*) $y = |\cos x|$ at the points $x = (k + \frac{1}{2})\pi$, $k = 0, \pm 1, \pm 2, \ldots$.

Find the derivatives of the following functions (**23** to **160**):

23. $x^5 - 4x^3 + 2x - 3$. **24.** $\frac{1}{4} - \frac{1}{3}x + x^2 - \frac{1}{2}x^4$.

25. $ax^2 + bx + c$. **26.** $-5x^3/a$. **27.** $at^m + bt^{m+n}$.

28. $ax^6 + b$. **29.** $\pi/x + \log 2$. **30.** $3x^{2/3} - 2x^{5/2} + x^{-3}$.

31. $x^2\sqrt[3]{x^2}$ **32.** $a/\sqrt[3]{x^2} - b/x\sqrt[3]{x}$.

33. $(a + bx)/(c + dx)$. **34.** $(2x + 3)/(x^2 - 5x + 5)$. **35.** $2/(2x - 1) - 1/x$.

36. $(1 + \sqrt{z})/(1 - \sqrt{z})$.

37. $5 \sin x + 3 \cos x$. **38.** $\tan x - \cot x$.

39. $(\sin x + \cos x)/(\sin x - \cos x)$. **40.** $2t \sin t - (t^2 - 2) \cos t$.

41. $\operatorname{arc} \tan x + \operatorname{arc} \cot x$. **42.** $x \cot x$. **43.** $x \operatorname{arc} \sin x$.

44. $\frac{1}{2}(1 + x^2) \operatorname{arc} \tan x - \frac{1}{2}x$ **45.** $x^7 e^x$.

46. $(x - 1) e^x$. **47.** $x^{-2}e^{-x}$. **48.** x^5e^{-x}. **49.** $e^x \cos x$.

50. $(x^2 - 2x + 2) e^x$. **51.** $e^x \operatorname{arc} \sin x$.

52. $x^2/\log x$. **53.** $x^3 \log x - \frac{1}{3}x^3$. **54.** $\frac{1}{x} + 2 \log x - (\log x)/x$.

55. $\left(\frac{ax + b}{c}\right)$. **56.** $(2a + 3by)^2$.

57. $(3 + 2x^2)^4$. **58.** $\frac{3}{56}(2x - 1)^{-7} - \frac{1}{24}(2x - 1)^{-6} - \frac{1}{40}(2x - 1)^{-5}$

59. $\sqrt{1 - x^2}$. **60.** $\sqrt[3]{a + bx^3}$

61. $(a^{2/3} - x^{2/3})^{3/2}$. **62.** $\tan x - \frac{1}{3}\tan^3 x + \frac{1}{5}\tan^5 x$.

63. $\sqrt{\cot x} - \sqrt{\cot a}$ **64.** $2x + 5 \cos^3 x$.

65. $\operatorname{cosec}^2 t + \sec^2 t$. **66.** $-\frac{1}{6}(1 - 3 \cos x)^{-2}$.

67. $\frac{1}{3}\cos^{-3} x - \cos^{-1} x$. **68.** $(\frac{3}{5} \sin x - \frac{2}{5} \cos x)^{1/2}$

69. $y = \sin^{2/3} x + \cos^{-3} x$. **70.** $\sqrt{1 + \operatorname{arc} \sin x}$

71. $\sqrt{\operatorname{arc} \tan x} - (\operatorname{arc} \sin x)^3$.

72. $1/\operatorname{arc} \tan x$. **73.** $\sqrt{x e^x + x}$. **74.** $\sqrt[3]{2e^x - 2^x + 1} + \log^5 x$.

75. $\sin^3 5x \cos^2 \frac{1}{3}x$.

76. $-\frac{11}{2}(x - 2)^{-2} - 4(x - 1)^{-1}$.

77. $-\frac{15}{4}(x - 3)^{-4} - \frac{10}{3}(x - 3)^{-3} - \frac{1}{2}(x - 3)^{-2}$.

78. $\frac{1}{8}x^8(1 - x^2)^{-4}$. **79.** $x^{-1}\sqrt{2x^2 - 2x + 1}$.

80. $\dfrac{x}{a^2\sqrt{a^2 + x^2}}$ **81.** $\dfrac{x^3}{3\sqrt{(1 + x^2)^3}}$

82. $\frac{3}{2}x^{2/3} + \frac{18}{7}x^{7/6} + \frac{9}{5}x^{5/3} + \frac{6}{13}x^{13/6}$

83. $\frac{1}{8}(1 + x^3)^{8/3} - \frac{1}{5}(1 + x^3)^{5/3}$. **84.** $\frac{4}{3}\left(\frac{x - 1}{x + 2}\right)^{1/4}$.

85. $x^4(a - 2x^3)^2$.

86. $\left(\dfrac{a + bx^n}{a - bx^n}\right)^m$

87. $\frac{9}{5}(x + 2)^{-5} - 3(x + 2)^{-4} + 2(x + 2)^{-3} - \frac{1}{2}(x + 2)^{-2}$.

88. $(a + x)\sqrt{a - x}$.

89. $\sqrt{(x + a)(x + b)(x + c)}$.

90. $\sqrt[3]{y + \sqrt{y}}$ **91.** $(2t + 1)(3t + 2)(3t + 2)^{1/3}$. **92.** $(2ay - y^2)^{-1/2}$.

93. $\log[\sqrt{1 + e^x} - 1] - \log[\sqrt{1 + e^x} + 1]$.

94. $\frac{1}{15}\cos^3 x\,(3\cos^2 x - 5)$.

95. $\dfrac{(\tan^2 x - 1)(\tan^4 x + 10\tan^2 x + 1)}{3\tan^3 x}$

96. $\tan^5 5x$. **97.** $\frac{1}{2}\sin(x^2)$

98. $\sin^2(t^3)$. **99.** $3\sin x\cos^2 x + \sin^3 x$.

100. $\frac{1}{3}\tan^3 x - \tan x + 1$ **101.** $-\dfrac{\cos x}{3\sin^3 x} + \frac{4}{3}\cot x$.

102. $\sqrt{a\sin^2 x + \beta\cos^2 x}$. **103.** $\arcsin(x^2) + \arccos(x^2)$.

104. $\frac{1}{2}(\arcsin x)^2\arccos x$. **105.** $\arcsin[(x^2 - 1)/x^2]$.

106. $\arcsin[x/\sqrt{1 + x^2}]$. **107.** $(\arccos x)/\sqrt{1 - x^2}$

108. $\dfrac{1}{\sqrt{b}}\arcsin\left(x\sqrt{\dfrac{b}{a}}\right)$ **109.** $\sqrt{a^2 - x^2} + a\arcsin(x/a)$.

110. $x\sqrt{a^2 - x^2} + a^2\arcsin(x/a)$. **111.** $\arcsin(1 - x) + \sqrt{2x - x^2}$.

112. $(x - \frac{1}{2})\arcsin\sqrt{x} + \frac{1}{2}\sqrt{x - x^2}$.

113. $\log(\arcsin 5x)$. **114.** $\arcsin(\log x)$.

115. $\arctan[x\sin a/(1 - x\cos a)]$.

116. $\frac{2}{3}\arctan\left(\frac{5}{3}\tan\dfrac{x}{2} + \frac{4}{3}\right)$.

117. $3b^2\arctan\sqrt{x/(b - x)} - (3b + 2x)\sqrt{bx - x^2}$.

118. $-\sqrt{2}\arctan(\tan x/\sqrt{2}) - x$. **119.** $\sqrt{e^{ax}}$. **120.** $e^{\sin^2 x}$

121. $(2ma^{mx} + b)^p$. **122.** $e^{at}\cos\beta t$.

123. $\dfrac{a\sin\beta x - \beta\cos\beta x}{a^2 + \beta^2}e^{ax}$. **124.** $\frac{1}{10}e^{-x}(3\sin 3x - \cos 3x)$.

125. $x^n a^{-x^2}$. **126.** $\sqrt{\cos x}\,a^{\sqrt{\cos x}}$. **127.** $3^{\cot(1/x)}$

128. $\log(ax^2 + bx + c)$. **129.** $\log[x + \sqrt{a^2 + x^2}]$.

130. $x - 2\sqrt{x} + 2\log(1 + \sqrt{x})$.

131. $\log[a + x + \sqrt{x^2 + 2ax}]$. **132.** $1/\log^2 x$.

133. $\log\cos\dfrac{x - 1}{x}$. **134.** $\log\dfrac{(x - 2)^5}{(x + 1)^3}$.

135. $\log\dfrac{(x - 1)^3(x - 2)}{x - 3}$. **136.** $-\frac{1}{2}\operatorname{cosec}^2 x + \log\tan x$.

137. $\frac{1}{2}x\sqrt{x^2-a^2}-\frac{1}{2}a^2\log\left[x+\sqrt{x^2-a^2}\right]$

138. $\log\log(3-2x^3)$. **139.** $5\log^3(ax+b)$.

140. $\log\dfrac{\sqrt{x^2+a^2}+x}{\sqrt{x^2+a^2}-x}$.

141. $\frac{1}{2}m\log(x^2-a^2)+\dfrac{n}{2a}\log\dfrac{x-a}{x+a}$.

142. $x\sin(\log x-\frac{1}{4}\pi)$. **143.** $\frac{1}{2}\log\tan\dfrac{x}{2}-\cos x/(2\sin^2 x)$.

144. $\sqrt{x^2+1}-\log[1+\sqrt{x^2+1}]+\log x$.

145. $\frac{1}{3}\log\dfrac{x^2-2x+1}{x^2+x+1}$ **146.** $2^{\text{arc sin } 3x}+(1-\text{arc cos } 3x)^2$.

147. $3^{\sin ax/\cos bx}+\frac{1}{3}\sin^3 ax/\cos^3 bx$.

148. $\dfrac{1}{\sqrt{3}}\log\dfrac{\tan\frac{1}{2}x+2-\sqrt{3}}{\tan\frac{1}{2}x+2+\sqrt{3}}$

149. $\text{arc tan}\log x$. **150.** $\log\text{arc sin } x+\frac{1}{2}\log^2 x+\text{arc sin}\log x$.

151. $\text{arc tan}\log\dfrac{1}{x}$. **152.** $\dfrac{\sqrt{2}}{3}\text{arc tan}\dfrac{x}{\sqrt{2}}+\dfrac{1}{6}\log\dfrac{x-1}{x+1}$.

153. $\log\dfrac{1+\sqrt{\sin x}}{1-\sqrt{\sin x}}$.

154. $\frac{3}{4}\log\dfrac{x^2+1}{x^2-1}+\frac{1}{4}\log\dfrac{x-1}{x+1}+\frac{1}{2}\text{arc tan } x$.

155. $\frac{1}{2}\log(1+x)-\frac{1}{6}\log(x^2-x+1)+\dfrac{1}{\sqrt{3}}\text{arc tan}\dfrac{2x-1}{\sqrt{3}}$.

156. $(x\,\text{arc sin})/\sqrt{1-x^2}+\log\sqrt{1-x^2}$.

157. $|x|$. **158.** $x|x|$. **159.** $\log|x|$ $(x\neq 0)$.

160. $f(x)=1-x$ if $x\leqslant 0$, $f(x)=e^{-x}$, $x>0$.

161. If $f(x)=e^{-x}$ find $f(0)+xf'(0)$.

162. If $f(x)=\sqrt{1+x}$ find $f(3)+(x-3)f'(3)$.

163. If $f(x)=\tan x$, $\Phi(x)=\log(1-x)$, find $f'(0)/\Phi'(0)$.

164. If $f(x)=1-x$, $\Phi(x)=1-\sin(\frac{1}{2}\pi x)$, find $\Phi'(1)/f'(1)$

165. Prove that the derivative of an even function is an odd function, but that the derivative of an odd function is an even function.

166. Prove that the derivative of a periodic function is a periodic function.

167. Prove that the function $y=xe^{-x}$ satisfies the equation $xy'=$ $=(1-x)y$.

168. Prove that the function $y=xe^{-x^2/2}$ satisfies the equation $xy'=$ $=(1-x^2)y$.

169. Prove that the function $y=(1+x+\log x)^{-1}$ satisfies the equation $xy'=y(y\log x-1)$.

By taking logarithmic derivatives find y' when y is given by (**170**—**179**):
170. $(x + 1)(2x + 1)(3x + 1)$. **171.** $(x + 2)^2 (x + 1)^{-3} (x + 3)^{-4}$.
172. $\sqrt{x(x - 1)/(x - 2)}$. **173.** $x[x^2/(x^2 + 1)]^{1/3}$.
174. $\sqrt[x]{x}$. **175.** $x^{\sqrt{x}}$. **176.** x^{x^x}. **177.** $x^{\sin x}$.
178. $(\cos x)^{\sin x}$. **179.** $(1 + x^{-1})^x$.

Find the derivative $y' = \mathrm{d}y/\mathrm{d}x$ when y is defined as a function of x by the parametric equations **180**—**192**:
180. $x = 2t - 1$, $y = t^3$. **181.** $x = (t + 1)^{-1}$, $y = t^2(t + 1)^{-2}$.
182. $x = 2at/(1 + t^2)$. $y = a(1 - t^2)/(1 + t^2)$.
183. $x = 3at/(1 + t^3)$, $y = 3at^2/(1 + t^3)$. **184.** $x = t^{1/2}$, $y = t^{1/3}$.
185. $x = \sqrt{1 + t^2}$, $y = (t - 1)/\sqrt{1 + t^2}$.
186. $x = a(\cos t + t \sin t)$, $y = a(\sin t - t \cos t)$.
187. $x = a \cos^2 t$, $y = b \sin^2 t$. **188.** $x = a \cos^3 t$, $y = b \sin^3 t$.
189. $x = \cos^3 t(\cos 2t)^{-1/2}$, $y = \sin^3 t(\cos 2t)^{-1/2}$.
190. $x = \operatorname{arc\,cos} (t^2 + 1)^{-1/2}$, $y = \operatorname{arc\,sin} t(t^2 + 1)^{-1/2}$.
191. $x = \mathrm{e}^{-t}$, $y = \mathrm{e}^{2t}$. **192.** $x = a(\log \tan \frac{1}{2} t + \cos t - \sin t)$, $y = a(\sin t + \cos t)$.
193. Find $\mathrm{d}y/\mathrm{d}x$ at $t = 1$ if $x = t \log t$, $y = t^{-1} \log t$.
194. Find $\mathrm{d}y/\mathrm{d}x$ at $t = \pi/4$ if $x = \mathrm{e}^t \cos t$, $y = \mathrm{e}^t \sin t$.
195. Prove that the function $y(x)$ defined by the parametric equations $x = 2t + 3t^2$, $y = t^2 + 2t^3$ satisfies the equation $y = (y')^2 + 2(y')^3$.
196. At the point $x=2$ we have the relation $x^2 = 2x$. Does $(x^2)'=(2x)'$ at $x = 2$?
197. Let $y = \sqrt{a^2 - x^2}$. Is it possible to differentiate the equation $x^2 + y^2 = a^2$ term by term?

In **198**—**215** find the derivative $y' = \mathrm{d}y/\mathrm{d}x$ for the implicit function y:
198. $2x - 5y + 10 = 0$. **199.** $x^2/a^2 + y^2/b^2 = 1$. **200.** $x^2 + y^2 = a^2$.
201. $x^3 + x^2y + y^2 = 0$. **202.** $\sqrt{x} + \sqrt{y} = \sqrt{a}$. **203.** $x^{2/3} + y^{2/3} = a^{2/3}$.
204. $(x + y) y^3 = x - y$. **205.** $y - 0.3 \sin y = x$.
206. $a \cos^2 (x + y) = b$. **207.** $\tan y = xy$.
208. $xy = \operatorname{arc\,tan}(x/y)$. **209.** $\operatorname{arc\,tan}(x + y) = x$.
210. $\mathrm{e}^y = x + y$. **211.** $\log x + \mathrm{e}^{-y/x} = c$. **212.** $\log y + x/y = c$.
213. $\operatorname{arc\,tan} (y/x) = \frac{1}{2} \log (x^2 + y^2)$.
214. $\sqrt{x^2 + y^2} = c \operatorname{arc\,tan} (y/x)$. **215.** $x^y = y^x$.
216. Find the angles at which the curves $y = \sin x$ and $y = \sin 2x$ cut the x-axis.
217. Find the angle at which the curve $y = \tan x$ cuts the x-axis.
218. Find the angle at which the curve $y = \mathrm{e}^{0.5 x}$ cuts the line $x = 2$.

219. Find the points at which the tangent to the curve $y = 3x^4 + 4x^3 - 12x^2 + 20$ is parallel to the x-axis.

220. Find the equation of the parabola $y = x^2 + bx + c$ which touches the line $y = x$ at the point $(1, 1)$.

221. Determine the slope of the tangent to the curve $x^3 + y^3 - xy - 7 = 0$ at the point $(1, 2)$.

222. At what point does the curve $y^2 = 2x^3$ have its tangent perpendicular to the straight line $4x - 3y + 2 = 0$?

223. Find the equations of the tangent and normal to the parabola $y = \sqrt{x}$ at the points with abscissa 4.

224. Write down the equations of the tangent and normal to the curve $y = x^3 + 2x^2 - 4x - 3$ at the point $(-2, 5)$.

225. Write down the equations of the tangent and normal to the curve $y = (x - 1)^{1/3}$ at the point $(1, 0)$.

226. Write down the equation of the tangent and normal at the point $(2, 2)$ to the curve $x = (1 + t)/t^3$, $y = 3t^{-2}/2 + t^{-1}/2$.

227. Write down the equations of the tangents to the curve $x = t \cos t$, $y = t \sin t$ at the points $t = 0$ and $t = \pi/4$.

228. Write down the equations of the tangent and normal to the curve $x^3 + y^2 + 2x - 6 = 0$ at the point with ordinate $y = 3$.

229. Write down the equation of the tangent to the curve $x^5 + y^5 = 2xy$ at the point $(1, 1)$.

230. Write down the equations of the tangents and normals at the points where the curve $y = (x - 1)(x - 2)(x - 3)$ cuts the x-axis.

231. Write down the equations of the tangent and normal to the curve $y^4 = 4x^4 + 6xy$ at the point $(1, 2)$.

232. Prove that the segment cut off by the coordinate axes of a tangent to the hyperbola $xy = a^2$ is bisected by the point of contact.

233. Prove that the segment cut off on the coordinate axes by a tangento tthe astroid $x^{2/3} + y^{2/3} = a^{2/3}$ has constant magnitude a.

234. Prove that the normals to the evolute of the curve $x = a(\cos t + t \sin t)$, $y = a(\sin t - t \cos t)$ are tangents to the circle $x^2 + y^2 = a^2$.

235. Find the angle at which the parabolas $y = (x - 2)^2$, $y = -4 + 6x - x^2$ intersect.

236. Find the angle of intersection of the parabolas $y = x^2$ and the curve $y = x^3$.

237. Prove that the hyperbolas $xy = a^2$, $x^2 - y^2 = b^2$ intersect at right angles.

238. For the parabola $y^2 = 4x$, calculate the lengths of the subtangent, the subnormal, the tangent and the normal at the point (1, 2).

239. Find the length of the subtangent at an arbitrary point on the curve $y = 2^x$.

240. Prove that the length of the normal at an arbitrary point on the rectangular hyperbola $x^2 - y^2 = a^2$ is equal to the length of the radius vector of that point.

241. Prove that the length of the subnormal at an arbitrary point on the hyperbola $x^2 - y^2 = a^2$ is equal to the abscissa of that point.

242. The motion of a point along the line OX is described by the equation $x = 3t - t^3$, x being measured in cm and t in sec. Find the velocity of the point at the times $t = 0, 1, 2$ sec.

243. The motion of two points along the x-axis is described by the equations $x = 100 + 5t$, and $x = t^2/2$ for $t \geqslant 0$. Find their relative velocity when one passes the other.

244. A rod AB of length 5 cm rests so that A is on the x-axis and B is on the y-axis and $OA = 3$ cm. If A begins to move along OX with velocity 2 cm/sec. find the initial velocity of B.

245. At time t a projectile fired with velocity v_0 in a direction making an angle a with the horizontal x-direction, is at (x, y) where $x = v_0 t \cos a$, $y = v_0 t \sin a - gt^2/2$. Find ($a$) the constraint equation of the path of the projectile; (b) the range of the projectile; (c) the magnitude and direction of the velocity at time t.

246. A point moves along the hyperbola $y = 10/x$ in such a way that its abscissa changes at a constant rate of 1 unit per second. At what rate is its ordinate changing at the point (5, 2)?

247. At what point on the parabola $y^2 = 18x$ is the ordinate changing twice as fast as the abscissa?

248. One side of a rectangle has fixed length $a = 10$ cm, while the other one b increases at the rate 4 cm/sec. At what rate is (a) the diagonal, (b) the area, of the rectangle increasing at the moment when $b = 30$ cm?

249. A point moves on the spiral $r = a\,\theta$ ($a = 10$ cm) in such a way that the angle θ changes at the rate of 6° per second. Find the rate at which r changes when it has the value 25 cm.

Find the second derivatives of the following functions **250**—**256**:

250. $x^8 + 7x^6 - 5x + 4$. **251.** e^{x^2}. **252.** $\sin^2 x$. **253.** $\log(1 + x^2)^{1/3}$.

254. $\log x + \sqrt{a^2 + x^2}$. **255.** $(1 + x^2) \arctan x$. **256.** $(\arcsin x)^2$.

257. Show that the function $y = \frac{1}{2} x^2 + x + 1$ satisfies the differential equation $2yy'' = 1 + y'^2$.

258. Show that the function $y = \frac{1}{2} x^2 e^x$ satisfies the differential equation $y'' - 2y' + y = e^x$.

259. Show that for any values of the constants c_1 and c_2, the function $y = c_1 e^{-x} + c_2 e^{-2x}$ satisfies the equation $y'' + 3y' + 2y = 0$.

260. Show that the function $y = e^{2x} \sin 5x$ satisfies the equation $y'' - 4y' + 29y = 0$.

261. If $y = x^3 - 5x^2 + 7x - 2$, find y'''.

262. If $f(x) = (2x - 3)^5$, find $f'''(3)$.

263. Find $y^{(v)}$ for the function $y = \log(1 + x)$.

264. Find $y^{(vi)}$ for the function $y = \sin 2x$.

265. Show that the function $y = e^{-x} \cos x$ satisfies the differential equation $y^{(iv)} + 4y = 0$.

266. If $f(x) = e^x \sin x$, find $f(0)$, $f'(0)$, $f''(0)$ and $f'''(0)$.

267. If the motion of a point along the x-axis is governed by the equation $x = 100 + 5t - 0.001t^3$, find the velocity and acceleration of the point at the instants $t = 0$, $t = 1$, $t = 10$.

268. Find the nth derivative of $(ax + b)^n$, n being a positive integer.

269. Find the nth derivative of the functions (*a*) $(1 - x)^{-1}$, (*b*) $x^{1/2}$.

270. Find the nth derivative of the functions: (*a*) $\sin x$; (*b*) $\cos 2x$; (*c*) e^{-3x}; (*d*) $\log(1 + x)$; (*e*) $(1 + x)^{-1}$; (*f*) $(1 + x)/(1 - x)$; (*g*) $\sin^2 x$; (*h*) $\log(ax + b)$.

271. Using Leibniz's formula find the nth derivative of the functions: (*a*) xe^x; (*b*) $x^2 e^{-2x}$; (*c*) $(1-x^2) \cos x$; (*d*) $y = (1 + x)/\sqrt{x}$; (*e*) $x^3 \log x$.

272. If $f(x) = \log [1/(1 - x)]$, find $f^{(n)}(0)$.

In Exercises **273—276** find d^2y/dx^2:

273. (*a*) $x = \log t$, $y = t^3$; (*b*) $x = \arctan t$, $y = \log(1 + t^2)$; (*c*) $x = \arcsin t$, $y = \sqrt{1 - t^2}$.

274. (*a*) $x = a \cos t$, $y = a \sin t$; (*b*) $x = a(t - \sin t)$, $y = a(1 - \cos t)$; (*c*) $x = a \cos^3 t$, $y = a \sin^3 t$; (*d*) $x = a(\sin t - t \cos t)$; $y = a(\cos t + t \sin t)$.

275. (*a*) $x = \cos 2t$, $y = \sin^2 t$; (*b*) $x = e^{-at}$, $y = e^{at}$.

276. $x = \arctan t$, $y = t^2/2$; (*b*) $x = \log t$, $y = (1 - t)^{-1}$.

277. Find d^2x/dy^2 if $x = e^t \cos t$, $y = e^t \sin t$.

278. Find d^2y/dx^2 for $t = 0$ if $x = \log(1 + t^2)$, $y = t^2$.

279. Show that if y is defined as a function of x through the parametric equations $x = \sin t$, $y = ae^{\sqrt{2}t} + be^{-\sqrt{2}t}$, where a and b are constants, it satisfies the differential equation $(1 - x^2)\, y'' - xy' = 2y$.

280. Find the increment Δy and the differential dy for the function $y = 5x + x^2$ if $x = 2$ and $\Delta x = 0.001$.

281. Without calculating the derivative find $d(1 - x^3)$ if $x = 1$ and $\Delta x = -1/3$.

282. The area S of a square with side of length x is given by the formula $S = x^2$. Find the increment and differential in S when x takes an increment Δx and give a geometrical interpretation of the results.

283. Give a geometrical interpretation of the increment and differential for the following functions: (*a*) area of a circle $S = \pi x^2$; (*b*) volume of a cube $v = x^3$.

284. Prove that as $\Delta x \to 0$ the increment in the function $y = 2^x$ corresponding to an increment of amount Δx in x is equal to $(2^x \log 2)\, \Delta x$.

285. For what value of x is the differential of the function $y = x^2$ not equivalent to the increment of the function as $\Delta x \to 0$?

286. Has the function $y = |x|$ a differential at $x = 0$?

287. Making use of the derivative, find the differential of the function $y = \cos x$ for $x = \pi/6$ and $\Delta x = \pi/36$.

289. Replacing the increment by the differential of the function, calculate approximately: (*a*) $\cos 61°$; (*b*) $\tan 44°$; (*c*) $e^{0 \cdot 2}$; (*d*) $\log 0.9$; (*e*) arc tan 1.05.

290. Find the approximate value of the increase in the volume of a sphere of radius 15 cm when its radius is increased by 2 mm.

291. Prove that, approximately,

$$\sqrt{x + \Delta x} = \sqrt{x} + \frac{\Delta x}{2\sqrt{x}},$$

and use this result to find approximate values of $\sqrt{5}$, $\sqrt{17}$, $\sqrt{70}$, $\sqrt{640}$.

292. Prove that, approximately,

$$\sqrt[3]{x + \Delta x} = \sqrt[3]{x} + \frac{\Delta x}{3\sqrt[3]{x^2}};$$

and find approximately the values of $\sqrt[3]{10}$, $\sqrt[3]{70}$, $\sqrt[3]{200}$.

293. Find approximate values of the functions: (*a*) $x^3 - 4x^2 + 5x + 3$ when $x = 1.03$; (*b*) $\sqrt{1 + x}$ when $x = 0.2$; (*c*) $\sqrt[3]{(1 - x)/(1 + x)}$ when $x = 0.1$; (*d*) $e^{1 - x^2}$ when $x = 1.05$.

294. Find an approximation to the value of $\tan 45°3'20''$.

295. Show that on the intervals $-1 \leqslant x \leqslant 0$ and $0 \leqslant x \leqslant 1$, the function $f(x) = x - x^3$ satisfies Rolle's theorem. Find the corresponding value of the mean value ξ in Rolle's formula.

296. The function $f(x) = \sqrt[3]{\overline{x-2}^2}$ assumes the equal values $f(0) = f(4) = \sqrt[3]{4}$ at the end-points of the interval $[0, 4]$. Does this function satisfy the conditions of Rolle's theorem in $[0, 4]$?

297. Does the function $f(x) = \tan x$ satisfy the conditions of Rolle's theorem in $[0, \pi]$?

298. If $f(x) = x(x+1)(x+2)(x+3)$, show that the equation $f'(x) = 0$ has three real roots.

299. The equation $e^x = 1 + x$ obviously has the root $x = 0$; show that it has no other real root.

Evaluate the limits **300—311**:

300. $\lim\limits_{x\to 0} \dfrac{x \cos x - \sin x}{x^3}$. **301.** $\lim\limits_{x\to 1} \dfrac{1-x}{(1-\sin\frac{1}{2}\pi x)}$.

302. $\lim\limits_{x\to 0} \dfrac{\tan x - \sin x}{x - \sin x}$. **303.** $\lim\limits_{x\to \frac{1}{4}\pi} \dfrac{\sec^2 x - 2\tan x}{1 + \cos 4x}$.

304. $\lim\limits_{x\to 1} (1-x) \tan\left(\frac{1}{2}\pi x\right)$. **305.** $\lim\limits_{x\to 0} \cot x \times \operatorname{arc\,sin} x$.

306. $\lim\limits_{x\to 0} x^n e^{-x}$, $(n > 0)$. **307.** $\lim\limits_{x\to\infty} x^n \sin(a/x)$, $n > 0$.

308. $\lim\limits_{x\to 1} \log x \times \log(x-1)$ **309.** $\lim\limits_{x\to 1} \dfrac{x}{x-1} + \dfrac{1}{\ln - x}$

310. $\lim\limits_{x\to 3} \left(\dfrac{1}{x-3} - \dfrac{5}{x^2-x-6}\right)$ **311.** (a) $\lim\limits_{\substack{x\to 0\\ y\to 0}} (x^2+y^2)\sin(1/xy)$;

(b) $\lim\limits_{\substack{x\to\infty\\ y\to\infty}} \dfrac{x+y}{x^2+y^2}$; (c) $\lim\limits_{\substack{x\to 0\\ y\to 2}} \dfrac{\sin xy}{x}$;

(d) $\lim\limits_{\substack{x\to\infty\\ y\to k}} \left(1+\dfrac{y}{x}\right)^x$; (e) $\lim\limits_{\substack{x\to 0\\ y\to 0}} \dfrac{x}{x+y}$; (f) $\lim\limits_{\substack{x\to 0\\ y\to 0}} \dfrac{x^2-y^2}{x^2+y^2}$.

312. Investigate the continuity of the function

$$f(x, y) = \begin{cases} \sqrt{1-x^2-y^2} & \text{if } x^2+y^2 \leqslant 1 \\ 0 & \text{if } x^2+y^2 > 1 \end{cases}$$

313. Find the points of discontinuity of the following functions: (a) $\log\sqrt{x^2+y^2}$; (b) $(x-y)^{-2}$; (c) $(1-x^2-y^2)^{-1}$; (d) $\cos(1/xy)$.

Find the first derivatives $\partial z/\partial x$ and $\partial z/\partial y$ of the functions **314—324**:

314. $z = x^3 + y^3 - 3axy$. **315.** $z = (x - y)/(x + y)$.

316. $z = y/x$. **317.** $z = \sqrt{x^2 - y^2}$.

318. $z = x/\sqrt{x^2 + y^2}$. **319.** $z = \log [x + \sqrt{x^2 + y^2}]$. **320.** $z = \text{arc } \tan(y/x)$

321. $z = x^y$. **322.** $z = \exp(\sin y/x)$.

323. $z = \text{arc} \sin \sqrt{(x^2 - y^2)/(x^2 + y^2)}$. **324.** $z = \log \sin[(x + a) y^{-1/2}]$.

325. If $u = x^{y^2}$, find the derivatives u_x, u_y.

326. If $u = \log(1 + xy)$ find u_x, u_y.

327. Find $\partial f/\partial x$, $\partial f/\partial y$ at $(2, 1)$ if $f(x, y) = \sqrt{xy + x/y}$.

328. Find $\partial f/\partial x$, $\partial f/\partial y$, at the point $(1, 2)$ for the function $f(x, y) = \log(xy)$.

329. For the function $f(x, y) = x^2y$ find the total increment Δf and the total differential $\mathrm{d}f$ at the point $(1, 2)$. Calculate $\Delta f - \mathrm{d}f$ if (*a*) $\Delta x = 1$, $\Delta y = 2$; (*b*) $\Delta x = 0.1$, $\Delta y = 0.2$.

330. For any pair of functions u, v, show that (*a*) $\mathrm{d}(u + v) = \mathrm{d}u + \mathrm{d}v$; (*b*) $\mathrm{d}(uv) = u\mathrm{d}v + v\mathrm{d}u$; (*c*) $\mathrm{d}(u/v) = (v\mathrm{d}u - u\mathrm{d}v)/v^2$.

Find the total differential of the following functions:

331. $z = x^3 + y^3 - 3xy$. **332.** $z = x^2y^3$. **333.** $z = (x^2 - y^2)/(x^2 + y^2)$.

334. $z = \sin^2 x + \cos^2 y$. **335.** $z = yx^y$. **336.** $z = \log(x^2 + y^2)$.

337. $f = \log(1 + x/y)$.

338. One side of a rectangle is $a = 10$ cm, the other is $b = 24$ cm. Find the approximate change in the length of the diagonal l (by calculating $\mathrm{d}l$) when a is increased by 4 mm and b is decreased by 1 mm. Calculate the exact value Δl.

339. Calculate approximately: (*a*) $(1.02)^3 \times (0.97)^2$; (*b*) $\sqrt{(4.05)^2 + (2.93)^2}$; (*c*) $\sin 32° \times \cos 59°$.

340. The period of a simple pendulum is given by the formula $T = 2\pi \sqrt{l/g}$. Find the change in T if there are changes α and β in l and g respectively.

341. Find $\mathrm{d}z/\mathrm{d}t$ if $z = x/y$ where $x = \mathrm{e}^t$, $y = \log t$.

342 Find $\mathrm{d}u/\mathrm{d}t$ if $u = \log \sin(xy^{-1/2})$ where $x = 3t^2$, $y = \sqrt{1 + t^2}$.

343. Find $\mathrm{d}u/\mathrm{d}t$ if $u = xyz$ where $x = 1 + t^2$, $y = \log t$, $z = \tan t$.

344. Find $\mathrm{d}u/\mathrm{d}t$ if $u = (x^2 + y^2)^{-1/2} z$, where $x = R \cos t$, $y = R \sin t$, $z = H$.

345. Find $\mathrm{d}z/\mathrm{d}x$ if $z = u^v$, where $u = \sin x$, $v = \cos x$.

CHAPTER III

INTEGRATION: THEORY AND APPLICATIONS

§ 8. Basic problems of the integral calculus. The indefinite integral

86. The concept of an indefinite integral. One of the basic tasks of the differential calculus is to find the derivative or differential of a given function.

The primary task of the integral calculus consists in the converse — finding the function, given its derivative or differential.

Let the derivative

$$y' = f(x)$$

or differential

$$dy = f(x)\,dx$$

be given of the unknown function y.

A function $F(x)$, possessing a given function $f(x)$ as its derivative, or $f(x)dx$ as its differential, is called a primitive of the given function $f(x)$.

If, for example,

$$f(x) = x^2,$$

a primitive of the function will be $F(x) = \frac{1}{3}x^3$; we have, in fact,

$$\left(\frac{1}{3}x^3\right)' = \frac{1}{3} \times 3x^2 = x^2.$$

Suppose that a primitive $F(x)$ of the given function $f(x)$ has been found, so that we have the relationship

$$F'(x) = f(x).$$

Since the derivative of an arbitrary constant C is equal to zero, we also have:

$$[F(x) + C]' = F'(x) = f(x),$$

i.e. the function $F(x) + C$ is also a primitive of $f(x)$.

Hence it follows that, if the problem of finding a primitive has one solution, it will have an infinity of further solutions, differing

from the first by an arbitrary constant. On the other hand, it can be shown that there are no other solutions apart from these, i.e.

If $F(x)$ is any one primitive of a given function $f(x)$, any other primitive has the form :

$$F(x) + C ,$$

where C is an arbitrary constant.

Let $F_1(x)$ be any function, whose derivative is $f(x)$. We have:

$$F_1'(x) = f(x) .$$

On the other hand, $F(x)$ possesses the derivative $f(x)$, i.e.

$$F'(x) = f(x) .$$

Subtracting this equation from the previous one, we get:

$$F_1'(x) - F'(x) = [F_1(x) - F(x)]' = 0 ,$$

whence, by the theorem proved in [63]:

$$F_1(x) - F(x) = C ,$$

where C is a constant: which it was required to prove.

The result we have obtained can also be formulated as: *If the derivatives (or differentials) of two functions are identically equal, the functions themselves differ only by a constant.*

The most general expression for a primitive is also referred to as the indefinite integral of the given function $f(x)$, or of the given differential $f(x)\mathrm{d}x$, and is denoted by the symbol

$$\int f(x)\mathrm{d}x ,$$

$f(x)$ being referred to as the integrand, and $f(x)\mathrm{d}x$ as the integrand expression.

Having found one primitive $F(x)$, we can write, by what was shown above:

$$\int f(x)\mathrm{d}x = F(x) + C ,$$

where C is an arbitrary constant.

Mechanical and geometrical interpretations can be given of the indefinite integral. Suppose we have a law giving an analytic relationship between velocity and time:

$$v = f(t) ,$$

and we want to express the path s in terms of time. Since the velocity of a point in a given trajectory is the derivative $\mathrm{d}s/\mathrm{d}t$ of the path

with respect to time, the problem reduces to finding a primitive of the function $f(t)$, i.e.

$$s = \int f(t)\, dt.$$

We get an infinite number of solutions, differing by a constant term. This lack of precision in the answer results from us not fixing the point from which the traversed path s is measured. If, for instance, $v = gt + v_0$ (uniformly accelerated motion), we obtain the expression for s:

$$s = \frac{1}{2} gt^2 + v_0 t + C, \qquad (1)$$

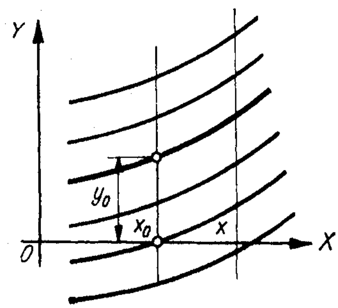

FIG. 114

because, as is easily shown, the derivative of (1) with respect to t coincides with the given expression $v = gt + v_0$. If we agree to measure s from the point corresponding to $t = 0$, i.e. if we agree to take $s = 0$ at $t = 0$, we have to put the constant $C = 0$ in (1). Of course it is of no significance that we have denoted the independent variable by t in the above discussion, and not by x.

We now pass to the geometrical interpretation of the problem of finding a primitive. The relationship $y' = f(x)$ shows that the graph of any required primitive, or, as we usually say, of any integral curve:

$$y = F(x),$$

is such that the tangent to the curve for any given x has the direction determined by the slope

$$y' = f(x). \qquad (2)$$

In other words, the direction of the tangent to the curve is given by (2) for any given value of the independent variable x; the problem is to find this curve. Having constructed one such integral curve, all the curves obtained by moving this by any amount in a direction parallel to the axis OY will have parallel tangents with the same slope $y' = f(x)$ as in the case of the initial curve, given the same value of x (Fig. 114). The parallel shift referred to is equivalent to adding a constant C to the ordinate of the curve; and the general equation of the curves, giving solutions of the problem, will be:

$$y = F(x) + C. \qquad (3)$$

In order to define fully the position of a curve, i.e. to fully define the expression for the primitive, a point must be assigned through which the integral curve must pass. The point assigned may be the point of intersection of the curve with a line

$$x = x_0$$

parallel to the axis OY. This is equivalent to assigning the initial value y_0 that the required function $y = F(x)$ takes for the given value $x = x_0$. We substitute this initial value in equation (3), and obtain an equation defining the arbitrary constant C:

$$y_0 = F(x_0) + C\,,$$

so that finally, the primitive satisfying our initial condition will have the form:

$$y = F(x) + [y_0 - F(x_0)].$$

Before examining the properties of the indefinite integral, and methods for finding the primitive, we note a second basic problem of the integral calculus, and examine it from the point of view of the problem already stated — namely, that of finding the primitive. A new concept is essential for what follows, this being the concept of a definite integral. We choose a natural approach here by starting from the intuitive idea of area, which also enables us to examine the connection between the concepts of definite integral and primitive. The discussion of the next two articles, based as it is on the intuitive idea of area, cannot be considered as a rigorous proof of new facts. A logically rigorous approach to the fundamentals of the integral calculus is indicated at the end of [88], whilst a full discussion is given at the end of the present chapter.

87. The definite integral as the limit of a sum. We take the graph of the function $f(x)$ in the plane XOY, and assume that it consists of a continuous curve, lying wholly above OX, i.e. all ordinates of the graph are assumed positive. We consider the area S_{ab} bounded by OX, the curve, and the two ordinates $x = a$ and $x = b$ (Fig. 115), and we try to find the magnitude of this area. We start by dividing the interval (a, b) into n parts by means of the points:

$$a = x_0 < x_1 < x_2 < \ldots < x_{k-1} < x_k < \ldots < x_{n-1} < x_n = b.$$

The area S_{ab} is now divided into n vertical strips, the length of base of the kth strip being $(x_k - x_{k-1})$. Let m_k and M_k respectively denote the least and greatest values of function $f(x)$ in the interval (x_{k-1}, x_k), i.e. the least and greatest ordinates of our graph in this interval. The area of the strip lies between the areas of the two rectangles of heights m_k and M_k and having the common base $(x_k - x_{k-1})$ (Fig. 116). These rectangles are the "interior" and "exterior" rectang-

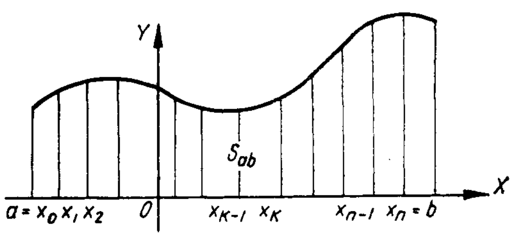

FIG. 115

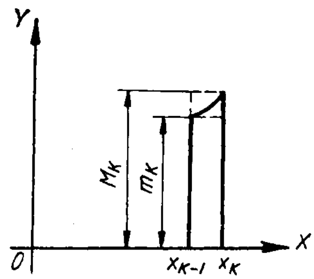

FIG. 116

les for the kth strip. The magnitude of the area of the kth strip is thus comprised between the areas of the rectangles in question, i.e. between the two numbers:

$$m_k(x_k - x_{k-1}) \text{ and } M_k(x_k - x_{k-1}),$$

and hence the total area S_{ab} will lie between the sums of the areas of these interior and exterior rectangles, i.e. S_{ab} lies between the sums:

$$\begin{aligned} s_n &= m_1(x_1 - x_0) + m_2(x_2 - x_1) + \ldots + m_k(x_k - x_{k-1}) + \ldots + \\ &\quad + m_{n-1}(x_{n-1} - x_{n-2}) + m_n(x_n - x_{n-1}), \qquad (4) \\ S_n &= M_1(x_1 - x_0) + M_2(x_2 - x_1) + \ldots + M_k(x_k - x_{k-1}) + \ldots + \\ &\quad + M_{n-1}(x_{n-1} - x_{n-2}) + M_n(x_n - x_{n-1}). \end{aligned}$$

We thus have the inequality:

$$s_n \leqslant S_{ab} \leqslant S_n. \qquad (5)$$

We now draw a mean rectangle in place of the interior and exterior rectangles for each strip, taking base $(x_k - x_{k-1})$ as usual but with the height taken as the ordinate $f(\xi_k)$ of our curve at any given point

ξ_k of the interval (x_{k-1}, x_k) (Fig. 117). We consider the sum of the areas of these *mean* rectangles:

$$\begin{aligned} S'_n = f(\xi_1)(x_1 - x_0) + f(\xi_2)(x_2 - x_1) + \ldots + f(\xi_k)(x_k - x_{k-1}) + \\ + \ldots + f(\xi_{n-1})(x_{n-1} - x_{n-2}) + f(\xi_n)(x_n - x_{n-1}). \end{aligned} \tag{6}$$

This, like the area S_{ab}, will lie between the sums of the areas of the interior and exterior rectangles, i.e. we have

$$s_n \leqslant S'_n \leqslant S_n. \tag{7}$$

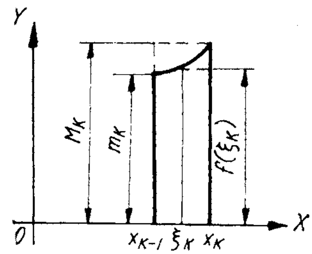

FIG. 117

We now indefinitely increase the number n of divisions of the interval (a, b), and so also make the greatest of the differences $(x_k - x_{k-1})$ tend to zero. Since $f(x)$ is continuous by hypothesis, the difference $(M_k - m_k)$ between its greatest and least values in the interval (x_{k-1}, x_k) will tend to zero with indefinite decrease of the length of this interval, irrespective of its position in the fundamental interval (a, b) (see the property of a continuous function in [35]). We thus have ε_n tending to zero on passing to the above-mentioned limit, where ε_n is the greatest of the differences:

$$(M_1 - m_1),\ (M_2 - m_2), \ldots, (M_k - m_k), \ldots,$$
$$(M_{n-1} - m_{n-1}),\ (M_n - m_n).$$

We now state the difference between the sums of the areas of the interior and exterior rectangles:

$$\begin{aligned} S_n - s_n = (M_1 - m_1)(x_1 - x_0) + (M_2 - m_2)(x_2 - x_1) + \ldots + \\ + (M_k - m_k)(x_k - x_{k-1}) + \ldots + (M_n - m_n)(x_n - x_{n-1}); \end{aligned}$$

on replacing all the $(M_k - m_k)$ by the greatest difference ε_n and recalling that all the $(x_k - x_{k-1})$ are positive, we have:

$$\begin{aligned} S_n - s_n \leqslant \varepsilon_n(x_1 - x_0) + \varepsilon_n(x_2 - x_1) + \ldots + \\ + \varepsilon_n(x_k - x_{k-1}) + \ldots + \varepsilon_n(x_n - x_{n-1}), \end{aligned}$$

i.e.

$$S_n - s_n \leqslant \varepsilon_n(x_n - x_0) = \varepsilon_n(b - a).$$

We can thus write:

$$0 \leqslant S_n - s_n \leqslant \varepsilon_n(b - a),$$

i.e.

$$\lim_{n \to \infty} (S_n - s_n) = 0. \tag{8}$$

On the other hand, we had for any n:

$$s_n \leqslant S_{ab} \leqslant S_n, \tag{9}$$

the magnitude of the area S_{ab} being a definite number. It follows directly from (8) and (9) that the magnitude of area S_{ab} is the common limit of s_n and S_n, i. e. of the areas of the interior and exterior rectangles:

$$\lim s_n = \lim S_n = S_{ab}.$$

But the sum S'_n of the mean rectangles lies between s_n and S_n, as we have seen, so that this must also tend to the area S_{ab}, i.e.

$$\lim S_n = S_{ab}.$$

The sum S'_n is more general than s_n or S_n, inasmuch as we can arbitrarily choose the ξ_k in the interval (x_{k-1}, x_k), and in particular, we can take $f(\xi_k)$ as equal to the least ordinate m_k or the greatest ordinate M_k.

With these choices, the sum S'_n transforms to s_n or S_n.

The above discussion leads us to the following:

Suppose the function $f(x)$ is continuous in an interval (a, b); suppose that, having divided the interval into n parts by the points

$$a = x_0 < x_1 < x_2 < \ldots < x_{k-1} < x_k < \ldots < x_{n-1} < x_n = b,$$

we find the corresponding value of the function $f(\xi_k)$ for any $x = \xi_k$ in the interval (x_{k-1}, x_k), and that we now form the sum:

$$\sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1}), \tag{10†}$$

then this sum tends to a definite limit on indefinite increase of the number n of divisions of the interval and on indefinite decrease of the greatest of the differences $(x_k - x_{k-1})$. This limit is equal to the area bounded by the axis OX, the graph of function $f(x)$, and the two ordinates $x = a$ and $x = b$.

† $\sum_{k=1}^{n} f(\xi_k)(x_k - x_{k-1})$ is an abbreviated notation for the sum (6).

The limit in question is referred to as the definite integral of the function $f(x)$ with respect to the variable x between the lower limit $x = a$ and the upper limit $x = b$; it is denoted by:

$$\int_a^b f(x)\,\mathrm{d}x.$$

We note that the existence of a limit I of the sum (10) in the case of indefinite decrease of the greatest of $(x_k - x_{k-1})$ amounts to the following assertion: for any given positive ε there exists a positive δ such that

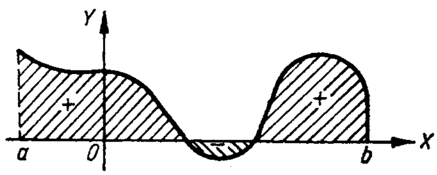

FIG. 118

$$\left| I - \sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1}) \right| < \varepsilon$$

for any chosen point ξ_k in the interval (x_{k-1}, x_k), provided only that all the (positive) differences $x_k - x_{k-1} < \delta$. This limit I is the definite integral.

We have assumed above that the graph of $f(x)$ is wholly located above axis OX, i.e. that all the ordinates of the graph are positive. We now take the general case, where part of the graph is above OX, and part below (Fig. 118).

If we form the sum (6) in this case, the terms $f(\xi_k)\,(x_k - x_{k-1})$ corresponding to parts of the graph below OX will be negative, since the difference $(x_k - x_{k-1})$ is positive and the ordinate $f(\xi_k)$ is negative.

The definite integral obtained on passing to the limit will reckon areas above OX with the (+) sign and areas below with the (−) sign, i.e. in the general case, *the definite integral*

$$\int_a^b f(x)\,\mathrm{d}x$$

will give the algebraic sum of the areas included between OX, the graph of $f(x)$, and the ordinates $x = a$ and $x = b$. Areas above OX are here given the positive sign, and areas below, the negative sign.

As will be seen later, finding the limit of a sum of the form (6) is not only involved in calculating areas, but is encountered in a wide variety of scientific problems. We take just one example. Let a certain point M be moving along axis OX from $x = a$ to $x = b$. Let it be acted on by a certain force T, also directed along OX. If the force T is constant, the work done in moving the point from the position $x = a$

to $x = b$ is given by the product $R = T\ (b - a)$, i.e. the product of the magnitude of the force and the path traversed by the point. If the force T is variable, the above formula is no longer valid. Suppose that the magnitude of the force depends on the position of the point on OX, i.e. we have $T = f(x)$.

To find the work done in this case, we subdivide the total path traversed by means of the points

$$a = x_0 < x_1 < x_2 < \ldots < x_{k-1} < x_k < \ldots < x_{n-1} < x_n = b,$$

and we take one of the intervals (x_{k-1}, x_k). We can take the force acting on the point as it moves from x_{k-1} to x_k as constant, with an error that is smaller, the shorter the length $(x_k - x_{k-1})$, and we can set its value as $f(\xi_k)$ for some point ξ_k of the interval (x_{k-1}, x_k). Hence we obtain an approximate expression for the work done in the interval (x_{k-1}, x_k):

$$R_k \sim f(\xi_k)\,(x_k - x_{k-1}).$$

The total work done will be given approximately by:

$$R \sim \sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1}).$$

On indefinite increase in the number n of subdivisions and on indefinite decrease of the greatest of the differences $(x_k - x_{k-1})$, we get in the limit a definite integral, accurately expressing the work done:

$$R = \int_a^b f(x)\,\mathrm{d}x.$$

Disregarding any possible geometrical or mechanical interpretations, we can now fix the concept of the definite integral of a function $f(x)$ over the interval $a \leqslant x \leqslant b$ as the limit of a sum of the form (6). The second basic task of the integral calculus is to study the properties of the definite integral and, above all, to evaluate it. If $f(x)$ is a given function, and $x = a$ and $x = b$ are given numbers, the definite integral

$$\int_a^b f(x)\,\mathrm{d}x.$$

is a determinate number. The $\int$ sign is a stylized letter S, recalling the summation that gives, in the limit, the magnitude of the definite integral. The expression under the integral, $f(x)\,\mathrm{d}x$, recalls the form of the term in the summation, viz., $f(\xi_k)\ (x_k - x_{k-1})$. The letter x, standing under the sign of the definite integral, is usually referred to as the *variable of integration*. We note an important detail as regards

this letter. The magnitude of the integral is a determinate number, as already mentioned, and is of course not dependent on the notation x for the variable of integration; any letter can be used to denote the variable of integration in a definite integral. The choice has evidently no influence at all on the magnitude of the integral, which depends only on the ordinates of the graph of $f(x)$ and on the limits of integration a and b. Since the notation for the independent variable plays no part, we have for instance:

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b f(t)\,\mathrm{d}t.$$

The second task of the integral calculus — that of evaluating the definite integral — consists in forming a sum of the form (6) and passing to the limit. This would seem a fairly complicated problem at first sight. We note that the number of terms in this sum increases indefinitely on passing to the limit, whilst each term tends to zero. Apart from this, the second task of the integral calculus would appear to have no connection with the first task, that of finding the primitive of a given function $f(x)$.

We show in the following article that both tasks are intimately related, and that evaluation of the definite integral $\int_a^b f(x)\,\mathrm{d}x$ is accomplished very simply, if the primitive of $f(x)$ is known.

88. The relation between the definite and indefinite integrals. We again consider the area S_{ab} bounded by the axis OX, the graph of function $f(x)$, and ordinates $x = a$ and $x = b$.

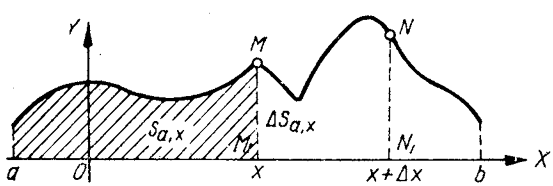

FIG. 119

In addition to this area, we also consider a part of it, bounded by the left-hand ordinate $x = a$ and by a movable ordinate, corresponding to a variable value of x (Fig. 119). The magnitude S_{ax} of this latter area will evidently depend on where we locate the right-hand ordinate, i.e. it is a function of x, represented by the definite integral of $f(x)$, taken from the lower limit a to the upper limit x. Since the letter x is

used to denote the upper limit, we shall avoid confusion by denoting the variable of integration by a second letter, say t. We can thus write:

$$S_{ax} = \int_a^x f(t)\,dt. \tag{11}$$

We have here a definite integral with a variable upper limit x, and its magnitude is clearly a function of this limit. We show that this function is one of the primitives of $f(x)$. We find the derivative of the function by considering its increment ΔS_{ax} for the increment Δx of the independent variable x. We obviously have (Fig. 120):

$$\Delta S_{ax} = \text{area } M_1MNN_1.$$

We denote the least and greatest ordinates of the graph of $f(x)$ in the interval $(x, x + \Delta x)$ by m and M respectively. The curvilinear figure M_1MNN_1, drawn on a larger scale in Fig. 120, will lie wholly inside the rectangle of height M and base Δx, and will contain the rectangle of height m and the same base; hence

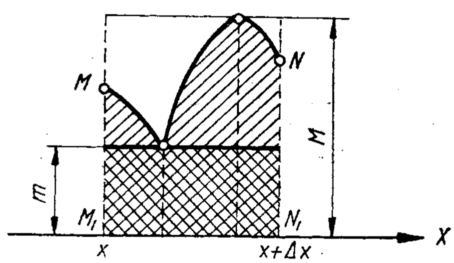

FIG. 120

$$m\Delta x \leqslant \Delta S_{xa} \leqslant M\Delta x\,,$$

or, dividing by Δx:

$$m \leqslant \frac{\Delta S_{ax}}{\Delta x} \leqslant M.$$

By the continuity of $f(x)$, both m and M tend to the common limit $M_1M = f(x)$, the ordinate of the curve at the point x, as Δx tends to zero; hence:

$$\lim \frac{\Delta S_{ax}}{\Delta x} = f(x)$$

which is what we wanted to prove. We can formulate the result obtained as follows: *a definite integral with a variable upper limit*

$$\int_a^x f(t)\,dt$$

is a function of this upper limit, the derivative of which is equal to the

integrand $f(x)$ *at the upper limit.* In other words, *the definite integral with variable upper limit is a primitive of the integrand.*

Having established the connection between the concepts of definite and indefinite integral, we now show how the definite integral

$$\int_a^b f(x)\,\mathrm{d}x$$

can be evaluated, if the primitive $F(x)$ of $f(x)$ is known. The definite integral with variable upper limit is in fact a primitive of $f(x)$, as we have shown, and we can write, by [86]:

$$\int_a^x f(t)\,\mathrm{d}t = F(x) + C, \tag{12}$$

where C is a constant. We determine this constant by noting that the area S_{ax} evidently vanishes if its right-hand ordinate coincides with the left-hand, i.e. if $x = a$; so that the left-hand side of (12) vanishes for $x = a$. Putting $x = a$ in (12) thus gives us:

$$0 = F(a) + C, \text{ i.e. } C = -F(a).$$

Substituting for C in (12), we get:

$$\int_a^x f(t)\,\mathrm{d}t = F(x) - F(a).$$

Finally, putting $x = b$ here, we find:

$$\int_a^b f(t)\,\mathrm{d}t = F(b) - F(a) \quad \text{or} \quad \int_a^b f(x)\,\mathrm{d}x = F(b) - F(a). \tag{13}$$

We thus arrive at the following fundamental rule, giving the magnitude of a definite integral in terms of values of a primitive: *the magnitude of a definite integral is equal to the difference between the values of the primitive of the integrand at the upper and lower limits of integration.*

The rule stated shows that finding a primitive, i.e. solving the first problem of the integral calculus, also solves the second problem, that of evaluating the definite integral; so that we do not need to carry out the complicated operations of forming the sum (6) and passing to the limit, in order to evaluate a definiteintegral.

We take as an example the definite integral

$$\int_0^1 x^2\,\mathrm{d}x.$$

A primitive of the function x^2 is $x^3/3$ [86].

We have by the rule we deduced:†

$$\int_0^1 x^2\,dx = \frac{1}{3}x^3\Big|_0^1 = \frac{1}{3}\times 1^3 - \frac{1}{3}\times 0^3 = \frac{1}{3}.$$

If we were to calculate this definite integral directly from its definition as the limit of a sum, without using the primitive, we should find ourselves with a much more complicated calculation, which is briefly reproduced. We divide the interval (0,1) into n equal parts with the points:

$$0 < \frac{1}{n} < \frac{2}{n} < \ldots < \frac{n-1}{n} < 1.$$

We now have the following n intervals:

$$\left(0,\ \frac{1}{n}\right),\ \left(\frac{1}{n},\ \frac{2}{n}\right),\ \left(\frac{2}{n},\ \frac{3}{n}\right)\ldots,\left(\frac{n-1}{n},\ 1\right),$$

the length of each being equal to $1/n$. We form the sum (6) with ξ_k taken as the left-hand end of the interval, i.e.

$$\xi_1 = 0,\quad \xi_2 = \frac{1}{n},\qquad \xi_3 = \frac{2}{n},\ldots,\xi_n = \frac{n-1}{n}.$$

We have $x_k - x_{k-1} = 1/n$ for all the differences, and we note that the integrand $f(x) = x^2$ has the values at the left-hand ends of the intervals:

$$f(\xi_1) = 0,\ f(\xi_2) = \frac{1}{n^2},\quad f(\xi_3) = \frac{2^2}{n^2},\ldots,\ f(\xi_n) = \frac{(n-1)^2}{n^2}.$$

So we can write:

$$\int_0^1 x^2\,dx = \lim_{n\to\infty}\left[0\times\frac{1}{n} + \frac{1}{n^2}\times\frac{1}{n} + \frac{2^2}{n^2}\times\frac{1}{n} + \ldots + \frac{(n-1)^2}{n^2}\times\right.$$
$$\left.\times\frac{1}{n}\right] = \lim_{n\to\infty}\frac{1^2 + 2^2\ldots + (n-1)^2}{n^3} \tag{14}$$

We find the sum in the numerator by noting the series of obvious equalities:

$$\begin{aligned}
(1+1)^3 &= 1 + 3\times 1 + 3\times 1^2 + 1^3\\
(1+2)^3 &= 1 + 3\times 2 + 3\times 2^2 + 2^3\\
(1+3)^3 &= 1 + 3\times 3 + 3\times 3^2 + 3^3\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
[1+(n-1)]^3 &= 1 + 3(n-1) + 3(n-1)^2 + (n-1)^3.
\end{aligned}$$

Adding term by term, we get:

$$2^3 + 3^3 + \ldots + n^3 = (n-1) + 3[1 + 2 + \ldots + (n-1)] +$$
$$+ 3[1^2 + 2^3 + \ldots + (n-1)^2] + 1^3 + 2^3 + \ldots + (n-1)^3.$$

† The symbol $\varphi(x)\Big|_a^b$ denotes the difference $[\varphi(b) - \varphi(a)]$.

Cancelling, and using the formula for the sum of an arithmetic progression, we can write:

$$n^3 = (n-1) + 3\,\frac{n(n-1)}{2} + 3\,[1^2 + 2^2 + 3^2 + \ldots + (n-1)^2] + 1 + 1,$$

whence

$$1^2 + 2^2 + 3^2 + \ldots + (n-1)^2 = \frac{n^3 - n}{3} - \frac{n(n-1)}{2} = \frac{n(n-1)\,(2n-1)}{6}\,.$$

Substituting the expression obtained in (14), we have:

$$\int_0^1 x^2\,dx = \lim_{n\to\infty} \frac{n(n-1)\,(2n-1)}{6n^3} =$$

$$= \frac{1}{6}\lim_{n\to\infty}\left(1-\frac{1}{n}\right)\left(2-\frac{1}{n}\right) = \frac{2}{6} = \frac{1}{3}\,.$$

We have now explained the basic problems of the integral calculus and their inter-relationship. We devote the next paragraph, to further consideration of the first problem, that of finding and investigating the properties of the indefinite integral.

Our above discussion of the definite integral has been based on purely geometrical concepts, viz., on consideration of the areas S_{ab} and S_{ax}. In particular, the proof of the basic fact that the sum (6) has a limit started from the assumption that there exists a definite area S_{ab} for every continuous curve. This assumption has no sound basis for all its apparent obviousness and the only mathematically rigorous course would be in the opposite direction: to prove the existence of a limit S of the sum

$$\sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1})$$

by direct analytic means, without regard to geometrical interpretation, then use the limit for defining the area S_{ab}. We give this proof at the end of the present chapter, and at the same time we make more general assumptions regarding $f(x)$ than those of continuity.

We also remark that a geometrical interpretation played an essential part in proving the basic proposition that, given continuity of the integrand, the derivative of the definite integral with respect to its upper limit is equal to the value of the integrand at the upper limit. A rigorous analytic proof of this proposition is given in the next section. Combining this proof with the proof of the existence of a definite integral of a continuous function, we are able to assert that

a primitive exists for every continuous function, i.e. an indefinite integral exists. We describe the basic properties of the indefinite integral later, on the assumption that we are only concerned with continuous functions.

We give a rigorous proof of the basic formula (13) when we come to describe the properties of definite integrals. Thus, the only unproved fact that remains is the existence of a limit of the sum (10) for a continuous function $f(x)$. This is proved at the end of the chapter, as already mentioned.

89. Properties of indefinite integrals. We saw in [86] that any two primitives of a given function can only differ by a constant term. This leads us to the first property of indefinite integrals:

I. *If two functions or two differentials are identical, their indefinite integrals can only differ by a constant term.*

Conversely, *to show that two functions differ by a constant term, it is sufficient to show that their derivatives (or differentials) are identical.*

The next properties, II and III, follow immediately from the concept of indefinite integral as a primitive, i.e. from the fact that *the indefinite integral*

$$\int f(x)\,\mathrm{d}x$$

is a function such that its derivative with respect to x is equal to the integrand $f(x)$, or that its differential is equal to the integrand expression $f(x)\,\mathrm{d}x$.

II. *The derivative of an indefinite integral is equal to the integrand, whilst its differential is equal to the integrand expression :*

$$\left(\int f(x)\,\mathrm{d}x\right)'_x = f(x);\ \ \mathrm{d}\int f(x)\,\mathrm{d}x = f(x)\,\mathrm{d}x. \tag{15}$$

III. We have, along with (15):

$$\int F'(x)\,\mathrm{d}x = F(x) + C,$$

and this can be rewritten [50] as:

$$\int \mathrm{d}F(x) = F(x) + C, \tag{16}$$

which, combined with property II, gives: *the signs* d *and* $\int$ *eliminate each other when juxtaposed in any order, provided we agree to neglect the arbitrary constant in the equation for an indefinite integral.*

IV. *A constant factor can be taken outside the integration sign :*

$$\int Af(x)\,\mathrm{d}x = A\int f(x)\,\mathrm{d}x + C.\dagger \tag{17}$$

V. *The integral of an algebraic sum is equal to the algebraic sum of the integrals of each term :*

$$\int (u + v - w + \ldots)\,\mathrm{d}x = \int u\mathrm{d}x + \int v\mathrm{d}x - \int w\mathrm{d}x + \ldots + C. \tag{18}$$

Formulae (17) and (18) are easily seen to be correct by differentiating both sides and observing the identity of the derivatives obtained. For (17), for instance:

$$(\int Af(x)\,\mathrm{d}x)' = Af(x);$$

$$(A\int f(x)\,\mathrm{d}x + C)' = A(\int f(x)\,\mathrm{d}x)' = Af(x).$$

90. Table of elementary integrals. This table is obtained by simply reading through the items of the table of elementary derivatives [49] in reverse order, when we get:

$$\int \mathrm{d}x = x + C.$$

$$\int x^m\,\mathrm{d}x = \frac{x^{m+1}}{m+1} + C, \quad \text{if} \quad m \neq -1.$$

$$\int \frac{\mathrm{d}x}{x} = \log x + C.$$

$$\int a^x\,\mathrm{d}x = \frac{a^x}{\log a} + C. \qquad \int \mathrm{e}^x\,\mathrm{d}x = \mathrm{e}^x + C.$$

$$\int \sin x\,\mathrm{d}x = -\cos x + C. \qquad \int \cos x\,\mathrm{d}x = \sin x + C.$$

$$\int \frac{\mathrm{d}x}{\cos^2 x} = \tan x + C. \qquad \int \frac{\mathrm{d}x}{\sin^2 x} = -\cot x + C.$$

$$\int \frac{\mathrm{d}x}{1+x^2} = \operatorname{arc\,tan} x + C. \qquad \int \frac{\mathrm{d}x}{\sqrt{1-x^2}} = \operatorname{arc\,sin} x + C.$$

To check this table, it is sufficient to establish that the derivative of the right-hand side of each equation is identical with the integrand on the left. In general, knowledge of the function, of which a given function $f(x)$ is the derivative, gives at once the indefi-

† The arbitrary constant is sometimes not written down after an indefinite integral, it being assumed that the indefinite integral already contains such a term. Equation (17) then runs:

$$\int Af(x)\mathrm{d}x = A\int f(x)\mathrm{d}x.$$

nite integral of the latter. Usually, however, even in the simplest examples, the given functions are not to be found in the table of derivatives, which makes problems of the integral calculus a good deal more difficult to work out than those of the differential calculus. It is always a case of transforming the given integral to one contained in the table of elementary integrals.

These transformations need experience and practice; they are facilitated by use of the basic rules of the integral calculus to be found below.

91. Integration by parts. If u, v are any two functions of x with continuous derivatives, we know that [50]:

$$\mathrm{d}(uv) = u\mathrm{d}v + v\mathrm{d}u, \text{ or } u\mathrm{d}v = \mathrm{d}(uv) - v\mathrm{d}u.$$

This gives us, using properties I, V and III:

$$\int u\mathrm{d}v = \int [\mathrm{d}(uv) - v\mathrm{d}u] + C = \int \mathrm{d}(uv) - \int v\mathrm{d}u + C =$$
$$= uv - \int v\mathrm{d}u + C,$$

leading to *the formula for integration by parts :*

$$\int u\mathrm{d}v = uv - \int v\mathrm{d}u + C. \tag{19}$$

We use this to pass from evaluating $\int u\mathrm{d}v$ to evaluating $\int v\mathrm{d}u$, the latter being possibly simpler.

Examples. **1.** $\int \log x \,\mathrm{d}x$.
Here we put

$$u = \log x, \quad \mathrm{d}x = \mathrm{d}v,$$

giving to start with:

$$\mathrm{d}u = \frac{\mathrm{d}x}{x}, \; v = x,$$

whence by (19):

$$\int \log x \,\mathrm{d}x = x \log x - \int x \frac{\mathrm{d}x}{x} + C = x \log x - x + C.$$

There is no need to write down the transformations separately in practice; as much as possible of the working should be done mentally.

2. $$\int \mathrm{e}^x x^2 \,\mathrm{d}x = \int x^2 \cdot \mathrm{e}^x \,\mathrm{d}x = \int x^2 \,\mathrm{d}\mathrm{e}^x = x^2 \mathrm{e}^x - \int \mathrm{e}^x \,\mathrm{d}x^2 =$$
$$= x^2 \mathrm{e}^x - 2 \int \mathrm{e}^x x\mathrm{d}x,$$
$$\int \mathrm{e}^x x\mathrm{d}x = \int x\mathrm{d}\mathrm{e}^x = x\mathrm{e}^x - \int \mathrm{e}^x \,\mathrm{d}x = \mathrm{e}^x x - \mathrm{e}^x,$$

which finally gives

$$\int e^x x^2 \, dx = e^x(x^2 - 2x + 2) + C.$$

3. $\int \sin x \cdot x^3 \, dx = \int x^3 \cdot \sin x \, dx = \int x^3 \, d(-\cos x) =$

$= -x^3 \cos x - \int (-\cos x) \, dx^3 = -x^3 \cos x + 3 \int x^2 \times \cos x \, dx =$

$= -x^3 \cos x + 3 \int x^2 \, d \sin x = -x^3 \cos x + 3x^2 \sin x - 3 \int \sin x \, dx^2 =$

$= -x^3 \cos x + 3x^2 \sin x - 6 \int x \sin x \, dx =$

$= -x^3 \cos x + 3x^2 \sin x - 6 \int x d(-\cos x) =$

$= -x^3 \cos x + 3x^2 \sin x + 6x \cos x - 6 \int \cos x \, dx =$

$= -x^3 \cos x + 3x^2 \sin x + 6x \cos x - 6 \sin x + C.$

The method indicated by these examples is used generally for evaluating integrals of the type:

$$\int \log x \cdot x^m dx, \ \int e^{ax} x^m dx, \ \int \sin bx \cdot x^m dx, \ \int \cos bx \cdot x^m dx,$$

where m is any positive integer; care is only needed to see that the power of x decreases with each successive transformation, until it reaches zero.

92. Rule for change of variables. Examples. An integral $\int f(x) dx$ can often be simplified by introducing a new variable t in place of x, putting

$$x = \varphi(t). \tag{20}$$

An indefinite integral can be transformed simply by substituting the new variable in the integrand expression :

$$\int f(x) \, dx = \int f[\varphi(t)] \, \varphi'(t) \, dt + C. \tag{21}$$

This is proved from property I of [89] by showing the identity of the differentials of the left and right-hand sides of (21). We have on differentiating:

$$d(\int f(x) \, dx) = f(x) \, dx = f[\varphi(t)] \, \varphi'(t) \, dt,$$

$$d(\int f[\varphi(t)] \, \varphi'(t) \, dt) = f[\varphi(t)] \, \varphi'(t) \, dt.$$

The inverse is often used instead of substitution (20):

$$t = \psi(x) \text{ and } \psi'(x) \, dx = dt.$$

Examples.

1. $\int (ax + b)^m \, dx$ (with $m \neq -1$).

The integral is simplified by substituting:

$$ax + b = t,\ a\mathrm{d}x = \mathrm{d}t,\ \mathrm{d}x = \frac{\mathrm{d}t}{a}.$$

Setting these values in the integral given, we find:

$$\int (ax+b)^m \,\mathrm{d}x = \frac{1}{a}\int t^m \,\mathrm{d}t = \frac{1}{a}\,\frac{t^{m+1}}{m+1} + C = \frac{1}{a}\,\frac{(ax+b)^{m+1}}{m+1} + C.$$

2. $$\int \frac{\mathrm{d}x}{ax+b} = \frac{1}{a}\int\frac{\mathrm{d}t}{t} = \frac{1}{a}\log t + C = \frac{\log\,(ax+b)}{a} + C.$$

3. $$\int\frac{\mathrm{d}x}{a^2+x^2} = \int\frac{\mathrm{d}x}{a^2\left(1+\dfrac{x^2}{a^2}\right)} = \frac{1}{a}\int\frac{\mathrm{d}\,(x/a)}{1+(x/a)^2} =$$

$$= \frac{1}{a}\,\mathrm{arc\ tan}\,\frac{x}{a} + C \left(\text{substituting } t = \frac{x}{a}\right).$$

4. $$\int\frac{\mathrm{d}x}{\sqrt{a^2-x^2}} = \int\frac{\mathrm{d}\,(x/a)}{\sqrt{1-(x/a)^2}} = \mathrm{arc\ sin}\,\frac{x}{a} + C.$$

5. $$\int\frac{\mathrm{d}x}{\sqrt{x^2+a}}.$$

We evaluate this integral by using Euler's substitution, about which more will be said below. We introduce the new variable t given by the formula:

$$\sqrt{x^2+a} = t - x, \quad t = x + \sqrt{x^2+a}.$$

We square both sides to find x and $\mathrm{d}x$:

$$x^2 + a = t^2 - 2tx + x^2,\ x = \frac{t^2-a}{2t} = \frac{1}{2}\left(t - \frac{a}{t}\right),$$

$$\sqrt{x^2+a} = t - \frac{t^2-a}{2t} = \frac{t^2+a}{2t},$$

$$\mathrm{d}x = \frac{1}{2}\left(1+\frac{a}{t^2}\right)\mathrm{d}t = \frac{1}{2}\,\frac{t^2+a}{t^2}\,\mathrm{d}t.$$

On carrying out all these substitutions in the integral given, we have:

$$\int\frac{\mathrm{d}x}{\sqrt{x^2+a}} = \int\frac{2t}{t^2+a}\times\frac{1}{2}\,\frac{t^2+a}{t^2}\,\mathrm{d}t = \int\frac{\mathrm{d}t}{t} =$$

$$= \log t + C = \log\left(x+\sqrt{x^2+a}\right) + C.$$

6. The integral

$$\int\frac{\mathrm{d}x}{x^2-a^2}$$

is evaluated by means of a special method, fully described later: that of *splitting the integrand into partial fractions.*

Having factorized the denominator of the integrand:

$$x^2 - a^2 = (x-a)\,(x+a)\,,$$

we write the integrand as a sum of simpler fractions:

$$\frac{1}{x^2 - a^2} = \frac{A}{x - a} + \frac{B}{x + a}.$$

Constants A and B are found by clearing fractions:

$$1 = A(x + a) + B(x - a) = (A + B)\, x + a(A - B)\,,$$

which must be true for any x. We thus find A and B from

$$a\,(A - B) = 1\,, \quad A + B = 0\,, \quad A = -B = \frac{1}{2a}.$$

We thus have:

$$\frac{1}{x^2 - a^2} = \frac{1}{2a}\left[\frac{1}{x - a} - \frac{1}{x + a}\right],$$

$$\int \frac{dx}{x^2 - a^2} = \frac{1}{2a}\left[\int \frac{dx}{x - a} - \int \frac{dx}{x + a}\right] =$$

$$= \frac{1}{2a}\,[\log\,(x - a) - \log\,(x + a)] + C = \frac{1}{2a}\log\frac{x - a}{x + a} + C.$$

7. Integrals of the more general type:

$$\int \frac{mx + n}{x^2 + px + q}\, dx$$

can be reduced to the forms already given by completing the square in the denominator of the integrand. We have:

$$x^2 + px + q = \left(x + \frac{1}{2}\,p\right)^2 + q - \frac{p^2}{4}.$$

We now put:

$$x + \frac{1}{2}\,p = t\,, \quad x = t - \frac{1}{2}\,p\,, \quad dx = dt\,,$$

giving

$$mx + n = m\left(t - \frac{1}{2}\,p\right) + n = At + B\,,$$

where

$$A = m \text{ and } B = n - \frac{1}{2}\,mp.$$

We finally put

$$q - \frac{p^2}{4} = \pm\, a^2\,,$$

where the $(+)$ or $(-)$ sign must be taken in accordance with the sign of the left-hand side of this equation, a being taken positive; so that we can rewrite the given integral as:

$$\int \frac{mx + n}{x^2 + px + q}\, dx = \int \frac{At + B}{t^2 \pm a^2}\, dt = A\int \frac{t\,dt}{t^2 \pm a^2} + B\int \frac{dt}{t^2 \pm a^2}.$$

The first of these integrals is evaluated at once by putting:

$$t^2 \pm a^2 = z; \quad 2t\mathrm{d}t = \mathrm{d}z\,,$$

which gives

$$\int \frac{t\mathrm{d}t}{t^2 \pm a^2} = \frac{1}{2}\int \frac{\mathrm{d}z}{z} = \frac{1}{2}\log z = \frac{1}{2}\log\,(t^2 \pm a^2).$$

The second integral has the form calculated in Example 3 (for $+$) or 6 (for $-$).

8. Integrals of the type

$$\int \frac{mx + n}{\sqrt{x^2 + px + q}}\,\mathrm{d}x$$

can be reduced to known forms by the same method of completing the square. Using the notation of Example 7, we can rewrite the integral as

$$\int \frac{mx + n}{\sqrt{x^2 + px + q}}\,\mathrm{d}x = \int \frac{At + B}{\sqrt{t^2 + b}}\,\mathrm{d}t =$$

$$= A\int \frac{t\mathrm{d}t}{\sqrt{t^2 + b}} + B\int \frac{\mathrm{d}t}{\sqrt{t^2 + b}} \left(b = \pm a^2 = q - \frac{p^2}{4}\right).$$

The first of these integrals is evaluated by substituting

$$t^2 + b = z^2\,, \quad 2t\mathrm{d}t = 2z\mathrm{d}z\,,$$

giving

$$\int \frac{t\mathrm{d}t}{\sqrt{t^2 + b}} = \int \frac{z\mathrm{d}z}{z} = \int \mathrm{d}z = z = \sqrt{t^2 + b}\,.$$

The second integral was worked out in Example 5 and is equal to $\log\left(t + \sqrt{t^2 + b}\right)$

9. A similar method of completing the square can be used to reduce

$$\int \frac{mx + n}{\sqrt{q + px - x^2}}\,\mathrm{d}x$$

to the form:

$$A_1\int \frac{t\mathrm{d}t}{\sqrt{a^2 - t^2}} + B_1\int \frac{\mathrm{d}t}{\sqrt{a^2 - t^2}}\,,$$

where we have

$$\int \frac{t\mathrm{d}t}{\sqrt{a^2 - t^2}} = -\sqrt{a^2 - t^2} + C$$

by using the substitution $a^2 - t^2 = z^2$. The second integral is worked out in Example 4.

10. $$\int \sin^2 x\,\mathrm{d}x = \int \frac{1 - \cos 2x}{2}\,\mathrm{d}x = \frac{1}{2}\left(x - \frac{1}{2}\sin 2x\right) + C =$$

$$= \frac{1}{2}(x - \sin x \cos x) + C.$$

$$\int \cos^2 x\,\mathrm{d}x = \int \frac{1 + \cos 2x}{2}\,\mathrm{d}x = \frac{1}{2}\left(x + \frac{1}{2}\sin 2x\right) + C =$$

$$= \frac{1}{2}(x + \sin x \cos x) + C.$$

11. The integral

$$\int \sqrt{x^2 + a}\, \mathrm{d}x$$

is reduced to a known form with the aid of integration by parts:

$$\int \sqrt{x^2 + a}\, \mathrm{d}x = x\sqrt{x^2 + a} - \int x . \mathrm{d}\sqrt{x^2 + a} =$$
$$= x\sqrt{x^2 + a} - \int \frac{x^2}{\sqrt{x^2 + a}}\, \mathrm{d}x .$$

Adding and subtracting a in the numerator of the last integrand, we can rewrite the above equation as

$$\int \sqrt{x^2 + a}\, \mathrm{d}x = x\sqrt{x^2 + a} - \int \sqrt{x^2 + a}\, \mathrm{d}x + a\int \frac{\mathrm{d}x}{\sqrt{x^2 + a}} ,$$

or

$$2\int \sqrt{x^2 + a}\, \mathrm{d}x = x\sqrt{x^2 + a} + a\int \frac{\mathrm{d}x}{\sqrt{x^2 + a}} ,$$

whence finally:

$$\int \sqrt{x^2 + a}\, \mathrm{d}x = \frac{1}{2}\left[x\sqrt{x^2 + a} + a \log\left(x + \sqrt{x^2 + a}\right)\right] + C .$$

93. Examples of differential equations of the first order. We considered some elementary differential equations in [51]. The general form of a first order differential equation is

$$F(x, y, y') = 0 .$$

This is a relationship connecting the independent variable x, the unknown function y and its first derivative y'. The equation can usually be solved with respect to y' and rewritten as:

$$y' = f(x, y) ,$$

where $f(x, y)$ is a known function of x and y.

We only pause to consider a few elementary examples, the general case of the equation being dealt with in Volume II.

Equation with separable variables. The function $f(x, y)$ is given here as the ratio of two functions, one of which only contains x, and the other only y:

$$y' = \frac{\varphi(x)}{\psi(y)} . \tag{22}$$

Since $y' = \mathrm{d}y/\mathrm{d}x$, we can rewrite this equation as:

$$\psi(y)\, \mathrm{d}y = \varphi(x)\, \mathrm{d}x ,$$

so that we have only x on one side, and only y on the other side: this is referred to as *separating the variables.* Since

$$\psi(y)\, \mathrm{d}y = \mathrm{d}\int \psi(y)\, \mathrm{d}y , \quad \varphi(x)\, \mathrm{d}x = \mathrm{d}\int \varphi(x)\, \mathrm{d}x ,$$

we have by property 1 [89]:

$$\int \psi(y)\,\mathrm{d}y = \int \varphi(x)\,\mathrm{d}x + C\,, \tag{23}$$

whence the required function y can be found by integration.

Examples. **1.** *Chemical reaction of the first order.* Denoting the amount of substance at the start of the reaction by a, and the amount taking part in the reaction at the instant t by x, we have the equation [51]:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = c\,(a - x)\,, \tag{24}$$

where c is a constant of the reaction. We also have the condition:

$$x\Big|_{t=0} = 0\,. \tag{25}$$

Separating the variables gives:

$$\frac{\mathrm{d}x}{a - x} = c\mathrm{d}t\,,$$

and integrating:

$$\int \frac{\mathrm{d}x}{a - x} = \int c\mathrm{d}t + C_1\,; \quad -\log\,(a - x) = ct + C_1\,,$$

where C_1 is an arbitrary constant. Hence,

$$a - x = \mathrm{e}^{-ct - C_1} = C\mathrm{e}^{-ct}\,,$$

where $C = \mathrm{e}^{-C_1}$ is also an arbitrary constant. The constant can be found by using condition (25), giving us $a = C$ for $t = 0$ from the last equation; so that finally:

$$x = a(1 - \mathrm{e}^{-ct})\,.$$

2. *Chemical reaction of the second order.* Suppose amounts a and b of two substances, measured in gram-molecules, are in solution at the start of the reaction. Suppose that equal quantities x, of the two substances, take part in the reaction at time t, so that the amounts remaining of the substances are $a - x$ and $b - x$.

By the basic law of chemical reactions of the second order, the reaction takes place with a speed proportional to the product of the remaining amounts, i.e.

$$\frac{\mathrm{d}x}{\mathrm{d}t} = k(a - x)\,(b - x)\,.$$

This equation has to be integrated, with the initial condition

$$x\,|_{t=0} = 0\,.$$

Separating the variables gives

$$\frac{\mathrm{d}x}{(a - x)\,(b - x)} = k\mathrm{d}t,$$

and integrating:

$$\int \frac{dx}{(a-x)(b-x)} = kt + C_1, \tag{26}$$

where C_1 is an arbitrary constant.

We use the method of partial fractions to evaluate the integral on the left (Example 6) [92]:

$$\frac{1}{(a-x)(b-x)} = \frac{A}{a-x} + \frac{B}{b-x},$$

$$1 = A(b-x) + B(a-x) = -(A+B)\,x + (Ab + Ba),$$

so that

$$-(A+B) = 0;\ Ab + Ba = 1,$$

whence

$$A = -B = \frac{1}{b-a}.$$

Hence

$$\int \frac{dx}{(a-x)(b-x)} = \frac{1}{b-a}\left[\int \frac{dx}{a-x} - \int \frac{dx}{b-x}\right]$$
$$= \frac{1}{b-a} \log \frac{b-x}{a-x}.$$

Substituting in (26), we have:

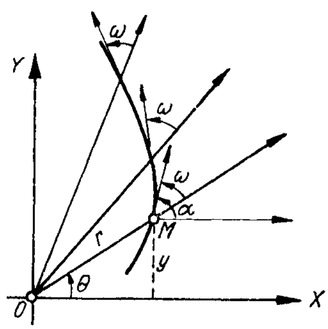

FIG. 121

$$\log \frac{b-x}{a-x} = (b-a)\,kt + (b-a)C_1,$$

$$\frac{b-x}{a-x} = Ce^{(b-a)kt},$$

where

$$C = e^{(b-a)C_1}.$$

The required function of x is now found without difficulty.

We propose that the reader work out the special case of $a = b$, when the above formula becomes meaningless.

3. *To find all the curves intercepting radius-vectors from the origin at a constant angle*† (Fig. 121).

Let $M(x, y)$ be a point of a required curve. We have from the figure:

$$\omega = \alpha - \theta,$$

$$\tan\omega = \tan(\alpha - \theta) = \frac{\tan\alpha - \tan\theta}{1 + \tan\alpha\tan\theta} = \frac{y' - \dfrac{y}{x}}{1 + y'\dfrac{y}{x}}.$$

† In general, the angle between two curves is defined as the angle between their tangents at the point of intersection.

Let $\tan\omega = \text{constant} = 1/a$; on multiplying up, we get the differential equation in the form:

$$x + yy' = a(y'x - y)\,,$$

or, multiplying both sides by $\mathrm{d}x$,

$$x\mathrm{d}x + y\mathrm{d}y = a(x\mathrm{d}y - y\mathrm{d}x)\,. \tag{27}$$

This equation is easily integrated, if we transform from rectangular coordinates x, y to polar coordinates r, θ, with OX as polar axis, and the origin O as pole. We have [82]:

$$x^2 + y^2 = r^2, \qquad \theta = \arctan\frac{y}{x}\,,$$

whence

$$x\mathrm{d}x + y\mathrm{d}y = r\mathrm{d}r, \quad \mathrm{d}\theta = \frac{1}{1+\dfrac{y^2}{x^2}}\,\mathrm{d}\frac{y}{x} = \frac{x\mathrm{d}y - y\mathrm{d}x}{x^2+y^2}\,.$$

Equation (27) now becomes:

$$r\mathrm{d}r = ar^2\,\mathrm{d}\theta \ \text{ i.e. } \ \frac{\mathrm{d}r}{r} = a\mathrm{d}\theta.$$

We get on integration:

$$\log r = a\theta + C_1,\ r = C\mathrm{e}^{a\theta}, \text{ where } C = \mathrm{e}^{C_1}.$$

The curves obtained are called *logarithmic spirals* [83].

§ 9. Properties of the definite integral

94. Basic properties of the definite integral. We have seen that the definite integral

$$\int_a^b f(x)\,\mathrm{d}x \tag{1}$$

is the limit of a sum of the form:

$$\sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1}) \quad (x_{k-1} \leqslant \xi_k \leqslant x_k). \tag{2}$$

We have assumed here that $a < b$ and correspondingly $x_{k-1} < x_k$

If we have $a > b$, integral (1) can be defined as before as the limit of the sum (2), except that now we have:

$$a = x_0 > x_1 > x_2 > \ldots > x_{k-1} > x_k > \ldots > x_{n-1} > x_n = b,$$

i.e. all the differences $x_k - x_{k-1}$ are negative. Finally, if we reverse limits a and b, i.e. we take a as upper limit and b as lower limit,

the points x_k of the interval must now be taken in the reverse order, whilst all $(x_k - x_{k-1})$ in (2) change sign, so that the sum itself and its limit change sign, i.e.

$$\int_b^a f(x)\,\mathrm{d}x = -\int_a^b f(x)\,\mathrm{d}x. \tag{3}$$

Further, on interpreting the definite integral as an area, it is natural to take

$$\int_a^a f(x)\,\mathrm{d}x = 0. \tag{4}$$

We also note the obvious equality:

$$\int_a^b \mathrm{d}x = b - a. \tag{5}$$

The function under the integral is here equal to unity for all x, so that

$$\int_a^b \mathrm{d}x = \lim\,[(x_1 - a) + (x_2 - x_1) + (x_3 - x_2) + \ldots + \\ + (x_{n-1} - x_{n-2}) + (b - x_{n-1})];$$

but the expression in square brackets is equal to the constant $(b - a)$. Evidently [87], (5) gives the area of a rectangle of base $(b - a)$ and unit height.

We can now start by noting three properties of definite integrals:

I. *The value of a definite integral with identical upper and lower limits is zero.*

II. *A definite integral preserves its absolute value and merely changes sign, on interchanging the upper and lower limits :*

$$\int_b^a f(x)\,\mathrm{d}x = -\int_a^b f(x)\,\mathrm{d}x.$$

For $a < b$, this property can be taken as defining the integral from b to a. It is naturally assumed that the integral on the right exists.

III. *The magnitude of a definite integral is independent of the notation for the variable of integration :*

$$\int_a^b f(x)\,\mathrm{d}x = \int_a^b f(t)\,\mathrm{d}t.$$

This has already been explained in [87].

The functions we consider will in future be assumed continuous in the interval of integration, unless there is a proviso to the contrary.

IV. *Given a series of numbers*

$$a, b, c, \ldots, k, l,$$

arranged in any order, we have:

$$\int_a^l f(x)\,dx = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx + \ldots + \int_k^l f(x)\,dx. \tag{6}$$

It suffices to establish this formula for the case of three numbers $a, b, c,$ the proof being then easily extended to cover any required number of terms.

We first take $a < b < c$. We have by definition:

$$\int_a^c f(x)\,dx = \lim \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}),$$

this limit being the same, irrespective of how we divide the interval (a, c), provided only that the greatest of the differences $(x_i - x_{i-1})$ tends to zero, and their number increases indefinitely. We can decide to divide (a, c) so that b, lying between a and c, appears as one of the points of division. The sum

$$\sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1})$$

can now be split into two sums of the same form, one being found by dividing the interval (a, b) and the other by dividing (b, c), with the number of divisions increasing indefinitely in both cases, and with the greatest of the $(x_i - x_{i-1})$ tending to zero. These two sums will tend respectively to

$$\int_a^b f(x)\,dx \text{ and } \int_b^c f(x)\,dx,$$

and we have finally:

$$\int_a^c f(x)\,dx = \lim \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}) = \int_a^b f(x)\,dx + \int_b^c f(x)\,dx,$$

which it was required to prove.

Now let b lie outside (a, c), say $a < c < b$. We can write in this case:

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx,$$

whence

$$\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x - \int_c^b f(x)\,\mathrm{d}x.$$

But by property II:

$$-\int_c^b f(x)\,\mathrm{d}x = \int_b^c f(x)\,dx,$$

i.e. we again have

$$\int_a^c f(x)\,\mathrm{d}x = \int_a^b f(x)\,\mathrm{d}x + \int_b^c f(x)\,\mathrm{d}x.$$

All the other possible arrangements of the points can be considered in a similar way.

V. *A constant factor can be taken outside the definite integration sign*, i.e.

$$\int_a^b Af(x)\,\mathrm{d}x = A\int_a^b f(x)\,\mathrm{d}x,$$

since

$$\int_a^b Af(x)\,\mathrm{d}x = \lim \sum_{i=1}^n Af(\xi_i)(x_i - x_{i-1}) =$$

$$= A \lim \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) = A\int_a^b f(x)\,\mathrm{d}x.$$

VI. *The definite integral of an algebraic sum is equal to the algebraic sum of the definite integrals of each term*, since, e.g.

$$\int_a^b [f(x) - \varphi(x)]\,\mathrm{d}x = \lim \sum_{i=1}^n [f(\xi_i) - \varphi(\xi_1)](x_i - x_{i-1}) =$$

$$= \lim \sum_{i=1}^n f(\xi_i)(x_i - x_{i-1}) - \lim \sum_{i=1}^n \varphi(\xi_i)(x_i - x_{i-1}) =$$

$$= \int_a^b f(x)\,\mathrm{d}x - \int_a^b \varphi(x)\,\mathrm{d}x.$$

95. Mean value theorem. VII. *If functions $f(x)$ and $\varphi(x)$ satisfy the condition*

$$f(x) \leqslant \varphi(x) \tag{7}$$

in the interval (a, b), then

$$\int_a^b f(x)\,\mathrm{d}x \leqslant \int_a^b \varphi(x)\,\mathrm{d}x \quad (b > a) \tag{8}$$

or briefly, *an inequality can be integrated.*

We form the difference

$$\int_a^b \varphi(x)\,\mathrm{d}x - \int_a^b f(x)\,\mathrm{d}x = \int_a^b [\varphi(x) - f(x)]\,\mathrm{d}x =$$

$$= \lim \sum_{i=1}^{n} [\varphi(\xi_i) - f(\xi_i)]\,(x_i - x_{i-1}).$$

By inequality (7), the terms under the summation are positive, or at least, not negative. The same can therefore be said of the whole sum and of its limit, i.e. of the difference between the integrals, whence follows inequality (8).

We also explain the above in geometrical terms. We first suppose that both curves

$$y = f(x),\ y = \varphi(x)$$

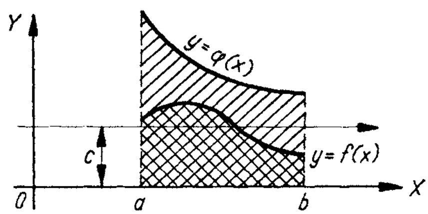

FIG. 122

lie above OX (Fig. 122). Then the figure bounded by the curve $y = f(x)$, OX and the ordinates $x = a$ and $x = b$, lies entirely within the similar figure bounded by $y = \varphi(x)$, and hence the area of the first figure cannot exceed that of the second figure, i.e.

$$\int_a^b f(x)\,\mathrm{d}x \leqslant \int_a^b \varphi(x)\,\mathrm{d}x.$$

The general case, with any arrangement of the given curves relative to OX, whilst preserving condition (7), follows from the above by giving the figure an upward displacement so that both curves appear above OX; this displacement adds the same term c to both functions $f(x)$ and $\varphi(x)$, and the same rectangular area with base $(b - a)$ and height c to the area of both figures, so that the inequality remains valid.

COROLLARY. *If we have in the interval* (a, b):

$$|f(x)| \leqslant \varphi(x) \leqslant M\,, \tag{9}$$

then

$$\left|\int_a^b f(x)\,\mathrm{d}x\right| \leqslant \int_a^b \varphi(x)\,\mathrm{d}x \leqslant M(b - a) \qquad (b > a). \tag{10}$$

In fact, (9) is equivalent to:

$$-M \leqslant -\varphi(x) \leqslant f(x) \leqslant \varphi(x) \leqslant M.$$

Integrating these inequalities from a to b (property VII) and using (5), we get:

$$-M(b-a) \leqslant -\int_a^b \varphi(x)\,dx \leqslant \int_a^b f(x)\,dx \leqslant \int_a^b \varphi(x)\,dx \leqslant M(b-a),$$

which is equivalent to inequality (10).

Setting $\varphi(x) = |f(x)|$, (10) gives the important inequality:

$$\left| \int_a^b f(x)\,dx \right| \leqslant \int_a^b |f(x)|\,dx, \tag{10_1}$$

which is a generalization for the case of an integral of the well-known property of a sum: the absolute magnitude of a sum is less than or equal to the sum of the absolute magnitudes of the component terms. It is easily seen that the sign of equality is obtained in the above formula only in the case when $f(x)$ does not change sign in the interval (a, b).

An extremely important theorem also follows from property VII.

Mean value theorem. *If the function $\varphi(x)$ preserves its sign in the interval (a, b), then*

$$\int_a^b f(x)\,\varphi(x)\,dx = f(\xi)\int_a^b \varphi(x)\,dx, \tag{11}$$

where ξ lies in the interval (a, b).

For clarity, we shall take $\varphi(x) \geqslant 0$ in the interval (a, b), whilst we denote the least and greatest values of $f(x)$ in (a, b) by m and M respectively. Since clearly

$$m \leqslant f(x) \leqslant M$$

(the signs of equality being obtained simultaneously only when $f(x)$ is constant), and $\varphi(x) \geqslant 0$, then

$$m\varphi(x) \leqslant f(x)\,\varphi(x) \leqslant M\varphi(x)\,,$$

and by property VII, taking $b > a$,

$$m\int_a^b \varphi(x)\,dx \leqslant \int_a^b f(x)\,\varphi(x)\,dx \leqslant M\int_a^b \varphi(x)\,dx.$$

Hence it is clear that there is a number P, satisfying $m \leqslant P \leqslant M$, such that

$$\int_a^b f(x)\,\varphi(x)\,dx = P\int_a^b \varphi(x)\,dx. \tag{12}$$

Since $f(x)$ is continuous, it takes in (a, b) all values included between m and M, one of these being P [35]. Hence there exists ξ in (a, b) such that

$$f(\xi) = P\,,$$

which proves formula (11).

If $\varphi(x) \leqslant 0$ in (a, b), then $-\varphi(x) \geqslant 0$ in (a, b). We get by applying the theorem just proved:

$$\int_a^b f(x)\,[-\varphi(x)]\,\mathrm{d}x = f(\xi)\int_a^b [-\varphi(x)]\,\mathrm{d}x;$$

on taking the $(-)$ sign outside the sign of integration and multiplying both sides by (-1), we arrive at formula (11).

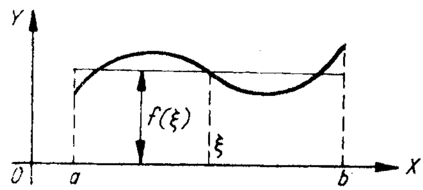

FIG. 123

If $b < a$, we have from the above in exactly the same way:

$$\int_b^a f(x)\,\varphi(x)\,\mathrm{d}x = f(\xi)\int_b^a \varphi(x)\,\mathrm{d}x.$$

Interchanging the limits of integration on both sides and multiplying by (-1), we arrive at (11), which is thus proved in general.

On putting $\varphi(x) = 1$, we obtain *an important particular case of the mean value theorem :*

$$\int_a^b f(x)\,\mathrm{d}x = f(\xi)\int_a^b \mathrm{d}x = f(\xi)\,(b-a). \tag{13}$$

The value of a definite integral is equal to the product of the length of the interval of integration and the value of the integrand for some value of the independent variable lying in the interval.

This length must be taken with the $(-)$ sign if $a > b$. This proposition means geometrically that, given the area bounded by any curve, OX and two ordinates $x = a$ and $x = b$, it is always possible to find a rectangle of the same area with the same base $(b - a)$ and with height *equal to some ordinate of the curve in the interval* (a, b) (Fig. 123).

It is easily shown that the ξ appearing in (11) and (13) can always be taken as lying inside (a, b).

96. Existence of the primitives. VIII. *If the upper limit of a definite integral is a variable, the derivative of the integral with respect to the upper limit is equal to the value of the integrand at the upper limit.*

We note that the value of

$$\int_a^b f(x)\,\mathrm{d}x$$

depends on the limits of integration a and b, given the integrand $f(x)$. We consider

$$\int_a^x f(t)\,\mathrm{d}t,$$

with constant lower limit a and variable upper limit x, the variable of integration being denoted by t to distinguish it from the upper limit x. The value of this integral will be a function of x:

$$F(x) = \int_a^x f(t)\,\mathrm{d}t. \tag{14}$$

We have to show that

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x).$$

We prove this by finding the derivative of $F(x)$ directly from the definition of [45]:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = \lim_{h\to 0} \frac{F(x+h) - F(x)}{h}.$$

We have:

$$F(x+h) = \int_a^{x+h} f(t)\,\mathrm{d}t = \int_a^x f(t)\,\mathrm{d}t + \int_x^{x+h} f(t)\,\mathrm{d}t$$

(by property IV), whence:

$$F(x+h) = F(x) + \int_x^{x+h} f(t)\,\mathrm{d}t,$$

and

$$\frac{F(x+h) - F(x)}{h} = \frac{1}{h}\int_x^{x+h} f(t)\,\mathrm{d}t.$$

Using (13), we have:

$$\int_x^{x+h} f(t)\,\mathrm{d}t = f(\xi)\cdot h,$$

where ξ lies in the interval $(x, x + h)$; hence

$$\frac{F(x+h) - F(x)}{h} = f(\xi).$$

As h tends to zero, ξ, lying between x and $x + h$, tends to x, and by the continuity of $f(x)$, $f(\xi)$ tends to $f(x)$, so that

$$\frac{dF(x)}{dx} = \lim_{h \to 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \to 0} f(\xi) = f(x),$$

which it was required to prove.

We note that h can only be given positive values for $x = a$, and only negative values for $x = b$ (with $a < b$), whilst $F(x)$ has the derivative $f(x)$ throughout (a, b) (closed). We have already discussed in [46] the definiton of the derivative at the ends of a closed interval.

The corollary follows [45] that *the definite integral $F(x)$, considered as a function of the upper limit x, is continuous in (a, b), where we must take $F(a) = 0$.*

We remark that if we apply the mean value theorem to integral (14), we get $F(x) = f(\xi)(x - a)$, whence it follows that $F(x) \to 0$ as $x \to a$. It also follows from the above discussion that:

IX. *Every continuous function $f(x)$ has a primitive or indefinite integral.*

Function (14) is the primitive of $f(x)$, which vanishes for $x = a$. If $F_1(x)$ is one primitive, then as we saw in [88]:

$$\int_a^b f(x)\,dx = F_1(b) - F_1(a). \tag{15}$$

97. Discontinuities of the integrand. It has been assumed in all the above discussions that the integrand $f(x)$ is continuous throughout the interval (a, b) of integration.

We now introduce the concept of integral for various discontinuous functions.

If the integrand $f(x)$ has a discontinuity at a point c of the interval (a, b), whilst each of the integrals

$$\int_a^{c-\varepsilon'} f(x)\,dx, \qquad \int_{c+\varepsilon''}^b f(x)\,dx \qquad (a < b)$$

tends to a definite limit as the positive numbers ε' and ε'' tend to zero,

these limits are referred to as the definite integrals of $f(x)$ between (a, c) and (c, b) respectively, i.e.

$$\int_a^c f(x)\,dx = \lim_{\varepsilon' \to +0} \int_a^{c-\varepsilon'} f(x)\,dx\,,$$

$$\int_c^b f(x)\,dx = \lim_{\varepsilon'' \to +0} \int_{c+\varepsilon''}^b f(x)\,dx\,,$$

if these limits exist.

We take in this case:

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

The function $F(x)$ defined by (14) is easily seen to have the following properties:

$F'(x) = f(x)$ for every point of (a, b) except $x = c$, and $F(x)$ is continuous throughout (a, b), excluding $x = c$.

If c coincides with one end of (a, b), only one of the limits has to be considered, either

$$\lim_{\varepsilon \to +0} \int_{a+\varepsilon}^b f(x)\,dx \quad \text{or} \quad \lim_{\varepsilon \to +0} \int_a^{b-\varepsilon} f(x)\,dx$$

Finally, if we have more than one point of discontinuity c in (a, b), the interval must be divided so that there is only one point of discontinuity in each interval.

Having agreed to attach the above meaning to the symbol

$$\int_a^b f(x)\,dx\,,$$

property IX *and formula* (15):

$$\int_a^b f(x)\,dx = F_1(b) - F_1(a)$$

will certainly be valid if $F'(x) = f(x)$ for every point of (a, b) except $x = c$, and $F_1(x)$ is continuous throughout (a, b), excluding $x = c$.

It is sufficient to prove this assertion for one discontinuity c inside (a, b), since the case of several discontinuities and the case of $c = a$ or $c = b$ follow in a similar way.

Since $f(x)$ is continuous in the intervals $(a, c - \varepsilon')$, $(c + \varepsilon'', b)$, we can apply (15) to these intervals, which gives:

$$\int_a^{c-\varepsilon'} f(x)\,dx = F_1(c - \varepsilon') - F_1(a)\,,$$

$$\int_{c+\varepsilon''}^{b} f(x)\,dx = F_1(b) - F_1(c + \varepsilon'')\,.$$

We can write, by the continuity of $F_1(x)$:

$$\int_a^c f(x)\,dx = \lim_{\varepsilon' \to +0} [F_1(c - \varepsilon') - F_1(a)] = F_1(c) - F_1(a)\,,$$

$$\int_c^b f(x)\,dx = \lim_{\varepsilon'' \to +0} [F_1(b) - F_1(c + \varepsilon'')] = F_1(b) - F_1(c)\,,$$

i.e.

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx =$$

$$= [F_1(c) - F_1(a)] + [F_1(b) - F_1(c)] = F_1(b) - F_1(a)\,,$$

which it was required to prove.

The case in question is encountered geometrically when a curve $y = f(x)$ has a discontinuity at a point c yet the the area under the curve always exists. Take, for instance, the graph of a function defined as follows:

$$f(x) = \frac{1}{2}x + \frac{1}{2} \quad \text{for} \quad 0 \leqslant x < 2\,,$$

$$f(x) = x \qquad \text{for} \quad 2 \leqslant x \leqslant 3$$

(Fig. 124).

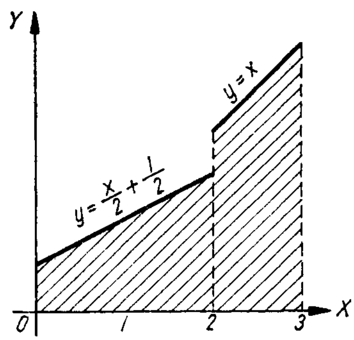

FIG. 124

The area bounded by this curve, OX, the ordinate $x = 0$ and the variable ordinate $x = x_1$, is a continuous function of x, in spite of the fact that $f(x)$ has a discontinuity at $x = 2$. On the other hand, it is easy to find a primitive of $f(x)$ that will be continuous throughout the interval (0,3). Take, for instance, the function $F_1(x)$ defined as follows:

$$F_1(x) = \frac{x^2}{4} + \frac{1}{2}x \quad \text{for} \quad 0 < x < 2\,,$$

$$F_1(x) = \frac{x^2}{2} \qquad \text{for} \quad 2 < x < 3\,.$$

We find, in fact, on differentiating:

$$F_1'(x) = \frac{1}{2}x + \frac{1}{2}$$

in (0,2), and $F_1'(x) = x$ in (2,3). Furthermore, the two expressions for $F_1(x)$ give the same value 2 for $x = 2$, which ensures the continuity of $F_1(x)$.

The area bounded by our curve, OX and the ordinates $x = 0$ and $x = 3$, is given by:

$$\int_0^3 f(x)\,dx = \int_0^2 f(x)\,dx + \int_2^3 f(x)\,dx = F_1(3) - F_1(0) = \frac{9}{2},$$

which is easily checked directly from the figure.

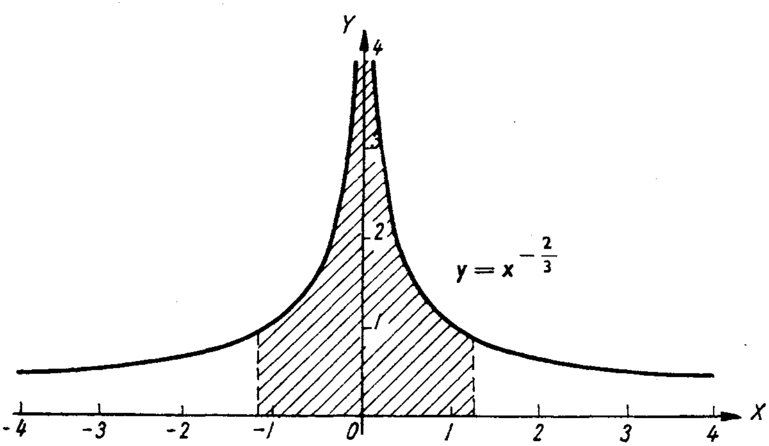

FIG. 125

We take further the function $y = x^{-2/3}$ (Fig. 125). It tends to infinity for $x = 0$, but its primitives remain continuous for this value of x, one primitive being $3x^{1/3}$; hence we can write:

$$\int_{-1}^{+1} x^{-2/3}\,dx = 3x^{1/3}\Big|_{-1}^{+1} = 6\,;$$

in other words, although the curve in question rises infinitely as x approaches zero, it still has a perfectly definite area between the ordinates $x = -1$ and $x = 1$.

The primitive $(-1/x)$ of the function $1/x^2$ itself tends to infinity for $x = 0$, so that (15) cannot be applied for this function in the case of zero lying inside the interval (a, b); the curve of $1/x^2$ does not possess a finite area for such an interval.

98. Infinite limits. The preceding discussion can be extended to the case of an *infinite interval*, on taking

$$\int_a^{+\infty} f(x)\,\mathrm{d}x = \lim_{b\to+\infty} \int_a^b f(x)\,\mathrm{d}x\,, \tag{16}$$

$$\int_{-\infty}^b f(x)\,\mathrm{d}x = \lim_{a\to-\infty} \int_a^b f(x)\,\mathrm{d}x. \tag{17}$$

provided these limits exist.

This proviso is certainly fulfilled if the primitive $F_1(x)$ tends to a definite limit as x tends to $(+\infty)$ or $(-\infty)$. Denoting these limits directly by $F_1(+\infty)$ and $F_1(-\infty)$, we have:

$$\int_a^{\infty} f(x)\,\mathrm{d}x = \lim_{b\to+\infty} [F_1(b) - F_1(a)] = F_1(+\infty) - F_1(a), \tag{18}$$

$$\int_{-\infty}^b f(x)\,\mathrm{d}x = \lim_{a\to-\infty} [F_1(b) - F_1(a)] = F_1(b) - F_1(-\infty), \tag{19}$$

$$\int_{-\infty}^{+\infty} f(x)\,\mathrm{d}x = \int_{-\infty}^{a} f(x)\,\mathrm{d}x + \int_a^{\infty} f(x)\,\mathrm{d}x = F_1(+\infty) - F_1(-\infty), \tag{20}$$

this last being a generalization of (15) for the case of an infinite interval.

Relationship (16) is often written as

$$\int_a^{\infty} f(x)\,\mathrm{d}x = \lim_{b\to+\infty} \int_a^b f(x)\,\mathrm{d}x.$$

Geometrically speaking, fulfilment of the above proviso means that the infinite branch of the curve $y = f(x)$, corresponding to $x \to \pm\infty$, has an area.

We have thus extended the concept of definite integral, originally established for a continuous function and a finite interval, to the cases of a discontinuous function and an infinite interval. This extension is characterized by first finding the integral of a continuous function for a shortened interval, then passing to the limit. Integrals obtained in this manner are distinguished from primitives by being referred to as *improper* integrals.

We note that the integral of a discontinuous function in a finite interval has in some cases a direct significance as the limit of a sum (cf. [94]). We shall discuss this later [116]. This is the case, e.g., with the integral expressing the area shown in Fig. 124. In essence, therefore, this integral will not be improper If, however, the integrand

is unbounded in the interval of integration (tends to infinity), or if this interval is infinite, the integral can then only exist in the improper form.

Example. The curve $y = 1/(1 + x^2)$ extends indefinitely for $x = \pm\infty$ yet always bounds a finite area with axis OX (Fig. 126), since

$$\int\limits_{-\infty}^{+\infty} \frac{dx}{1 + x^2} = \text{arc tan } x \Big|_{-\infty}^{+\infty} = \frac{1}{2}\pi - \left(-\frac{1}{2}\pi\right) = \pi.$$

It should be recalled, when evaluating this integral, that an arbitrary value of the many-valued function arc tan x cannot be taken; the function

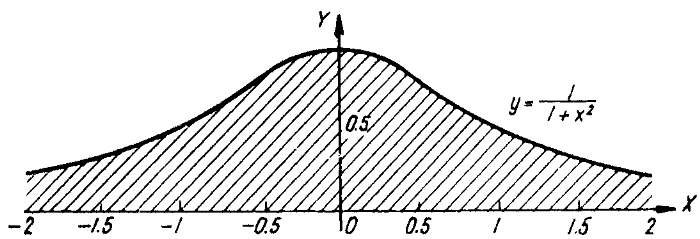

FIG. 126

must be defined as in [24] so that it becomes single-valued, i.e. its values lie between $(-\pi/2)$ and $(+\pi/2)$; if this is not done, the above formula becomes meaningless.

99. Change of variable for definite integrals. Let $f(x)$ be continuous in the interval (a, b), or in the wider interval (A, B), discussed below. Further, let $\varphi(t)$ be a single-valued, continuous function with a continuous derivative $\varphi'(t)$ in the interval (α, β), where

$$\varphi(\alpha) = a \text{ and } \varphi(\beta) = b. \tag{21}$$

We further suppose that $\varphi(t)$ does not move outside (a, b), or the wider interval (A, B), in which $f(x)$ is continuous, when t varies in (α, β). The function of a function $f[\varphi(t)]$ is now a continuous function of t in the interval (α, β).

If, with these assumptions, we introduce a new variable of integration t in place of x:

$$x = \varphi(t)\ , \tag{22}$$

the formula for transforming the definite integral is:

$$\int\limits_a^b f(x)\,dx = \int\limits_\alpha^\beta f[\varphi(t)]\,\varphi'(t)\,dt. \tag{23}$$

We prove this by introducing integrals with variable limits in place of those above:

$$F(x) = \int_a^x f(y)\,\mathrm{d}y; \qquad \Psi(t) = \int_a^t f[\varphi(z)]\,\varphi'(z)\,\mathrm{d}z.$$

By (22), $F(x)$ is a function of a function of t:

$$F(x) = F[\varphi(t)] = \int_a^{\varphi(t)} f(y)\,\mathrm{d}y.$$

We find its derivative by the rule for differentiation of a function of a function:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}t} = \frac{\mathrm{d}F(x)}{\mathrm{d}x}\,\frac{\mathrm{d}x}{\mathrm{d}t},$$

whilst by property VIII [96]:

$$\frac{\mathrm{d}F(x)}{\mathrm{d}x} = f(x);$$

it also follows from (22) that

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \varphi(t),$$

whence

$$\frac{\mathrm{d}F(x)}{\mathrm{d}t} = f(x)\,\varphi'(t) = f[\varphi(t)]\,\varphi'(t).$$

We now find the derivative of $\Psi(t)$. We have by property VIII and the assumptions made:

$$\frac{\mathrm{d}\Psi(t)}{\mathrm{d}t} = f[\varphi(t)]\,\varphi'(t).$$

Functions $\Psi(t)$ and $F(x)$, considered as functions of t, thus have the same derivative in the interval (α, β), and hence [89] can only differ by a constant; whilst we have for $t = \alpha$:

$$x = \varphi(\alpha) = a, \; F(x)\,|_{t=\alpha} = F(a) = 0; \; \Psi(\alpha) = 0,$$

i.e. these two functions are equal for $t = \alpha$ and hence are equal for all t in (α, β). In particular, we have for $t = \beta$:

$$F(x)\,|_{t=\beta} = F(b) = \int_a^b f(x)\,\mathrm{d}x = \int_\alpha^\beta f[\varphi(t)]\,\varphi'(t)\,\mathrm{d}t,$$

which it was required to prove.

The inverse:

$$t = \psi(x) \tag{24}$$

is very often used instead of substitution (22):

$$x = \varphi(t).$$

Limits α and β are then immediately defined by:

$$\alpha = \psi(a), \ \beta = \psi(b),$$

whilst it must be borne in mind that *expression* (22) *for* x, *obtained by solving* (24) *with respect to* x, *must satisfy all the conditions mentioned above ;* in particular, $\varphi(t)$ must be a *single-valued* function of t. It can be shown that (23) is invalid if $\varphi(t)$ lacks this property.

We replace x in $\int_{-1}^{+1} dx = 2$ by the new variable t, where $t = x^2$, and obtain an integral equal to zero from the right-hand side of (23), since its limits are the same, $+1$; but this is impossible; the error arises due to the expression for x in terms of t:

$$x = \pm\sqrt{t}$$

being a many-valued function.

Example. We call $f(x)$ an even function of x if $f(-x) = f(x)$, and an odd function if $f(-x) = -f(x)$.

For instance, $\cos x$ is an even function of x, and $\sin x$ an odd function.

We show that

$$\int_{-a}^{+a} f(x)\,dx = 2\int_0^a f(x)\,dx,$$

if $f(x)$ is even, and that

$$\int_{-a}^{+a} f(x)\,dx = 0,$$

if $f(x)$ is odd.

We separate the integral into two parts [94, IV]:

$$\int_{-a}^{+a} f(x)\,dx = \int_{-a}^{0} f(x)\,dx + \int_0^a f(x)\,dx.$$

We make a change of variable, $x = -t$, in the first integral and use properties II and III [94]:

$$\int_a^0 f(x)\,dx = -\int_a^0 f(-t)\,dt = \int_0^a f(-t)\,dt = \int_0^a f(-x)\,dx,$$

whence, substituting in the previous equation:

$$\int_{-a}^{+a} f(x)\,dx = \int_0^a f(-x)\,dx + \int_0^a f(x)\,dx = \int_0^a [f(-x) + f(x)]\,dx.$$

If $f(x)$ is an even function, $[f(-x) + f(x)] = 2f(x)$, whilst the sum in the square brackets is zero if $f(x)$ is odd; and this proves our statement.

100. Integration by parts. *In the case of definite integrals, the formula for integration by parts* [91] *can be put in the form :*

$$\int_a^b u(x)\,\mathrm{d}v(x) = u(x)\,v(x)\Big|_a^b - \int_a^b v(x)\,\mathrm{d}u(x). \tag{25}$$

We integrate term by term the identity [91]:

$$u(x)\,\mathrm{d}v(x) = \mathrm{d}[u(x)\,v(x)] - v(x)\,\mathrm{d}u(x)\,,$$

and obtain, in fact:

$$\int_a^b u(x)\,\mathrm{d}v(x) = \int_a^b \mathrm{d}[u(x)\,v(x)] - \int_a^b v(x)\,\mathrm{d}u(x),$$

whilst by property IX [96]:

$$\int_a^b \mathrm{d}[u(x)\,v(x)] = \int_a^b \frac{\mathrm{d}[u(x)\,v(x)]}{\mathrm{d}x}\,\mathrm{d}x = u(x)\,v(x)\Big|_a^b,$$

and hence we get (25). Of course it is assumed that $u(x)$ and $v(x)$ have continuous derivatives in the interval (a, b).

Example. To evaluate:

$$\int_0^{\frac{1}{2}\pi} \sin^n x\,\mathrm{d}x, \quad \int_0^{\frac{1}{2}\pi} \cos^n x\,\mathrm{d}x.$$

We put

$$I_n = \int_0^{\frac{1}{2}\pi} \sin^n x\,\mathrm{d}x.$$

Integrating by parts, we have:

$$I_n = \int_0^{\frac{1}{2}\pi} \sin^{n-1} x \sin x\,\mathrm{d}x = -\int_0^{\frac{1}{2}\pi} \sin^{n-1} x\,\mathrm{d}\cos x =$$

$$= -\sin^{n-1} x \cos x\Big|_0^{\frac{1}{2}\pi} + \int_0^{\frac{1}{2}\pi} (n-1)\sin^{n-2} x \cos x \cdot \cos x\,\mathrm{d}x =$$

$$= (n-1)\int_0^{\frac{1}{2}\pi} \sin^{n-2} x \cos^2 x\,\mathrm{d}x = (n-1)\int_0^{\frac{1}{2}\pi} \sin^{n-2} x(1-\sin^2 x)\,\mathrm{d}x =$$

$$= (n-1)\int_0^{\frac{1}{2}\pi} \sin^{n-2} x\,\mathrm{d}x - (n-1)\int_0^{\frac{1}{2}\pi} \sin^n x\,\mathrm{d}x =$$

$$= (n-1)\,I_{n-2} - (n-1)\,I_n,$$

i.e.

$$I_n = (n-1)\,I_{n-2} - (n-1)\,I_n\,,$$

whence, solving for I_n:

$$I_n = \frac{n-1}{n}\,I_{n-2}. \tag{26}$$

This is called a *reduction formula,* since it reduces the evaluation of I_n to the evaluation of a similar integral, but with a lower subscript $(n-2)$.

We take the cases separately, of n even or odd.

1. $n = 2k$ (even). We have by (26):

$$I_{2k} = \frac{2k-1}{2k}\,I_{2k-2} = \frac{(2k-1)\,(2k-3)}{2k(2k-2)}\,I_{2k-4} = \dots$$
$$= \frac{(2k-1)\,(2k-3)\dots 3\times 1}{2k(2k-2)\dots 4\cdot 2}\,I_0$$

and since

$$I_0 = \int_0^{\frac{1}{2}\pi} \mathrm{d}x = \frac{1}{2}\pi,$$

we have finally:

$$I_{2k} = \frac{(2k-1)\,(2k-3)\dots 3\times 1}{2k(2k-2)\dots 4\times 2}\cdot\frac{\pi}{2}\,.$$

2. $n = 2k+1$ (odd). We find, as above:

$$I_{2k+1} = \frac{2k(2k-2)\dots 4\times 2}{(2k+1)\,(3k-1)\dots 5\times 3}\,I_1, \quad I_1 = \int_0^{\frac{1}{2}\pi} \sin x\,\mathrm{d}x =$$
$$= -\cos x\,\Big|_0^{\frac{1}{2}\pi} = 1,$$

and hence

$$I_{2k+1} = \frac{2k(2k-2)\dots 4\times 2}{(2k+1)\,(2k-1)\dots 5\times 3}\,.$$

The integral

$$\int_0^{\frac{1}{2}\pi} \cos^n x\,\mathrm{d}x$$

can be evaluated in a similar way, though it is easier to reduce it to the former by noting that

$$\int_0^{\frac{1}{2}\pi} \cos^n x\,\mathrm{d}x = \int_0^{\frac{1}{2}\pi} \sin^n\left(\frac{1}{2}\pi - x\right)\mathrm{d}x,$$

whence, putting

$$\frac{1}{2}\pi - x = t, \quad x = \frac{1}{2}\pi - t,$$

we have on the basis of (23) and property II [94]:

$$\int_0^{\frac{1}{2}\pi} \cos^n x \, \mathrm{d}x = -\int_{\frac{1}{2}\pi}^{0} \sin^n t \, \mathrm{d}t = \int_0^{\frac{1}{2}\pi} \sin^n t \, \mathrm{d}t.$$

Taking together the results obtained, we can write:

$$\int_0^{\frac{1}{2}\pi} \sin^{2k} x \, \mathrm{d}x = \int_0^{\frac{1}{2}\pi} \cos^{2k} x \, \mathrm{d}x = \frac{(2k-1)(2k-3)..3\times 1}{2k(2k-2)..4\cdot 2} \, \frac{\pi}{2}, \tag{27}$$

$$\int_0^{\frac{1}{2}\pi} \sin^{2k+1} x \, \mathrm{d}x = \int_0^{\frac{1}{2}\pi} \cos^{2k+1} x \, \mathrm{d}x = \frac{2k(k-2)..4\times 2}{(2k+1)(2k-1)..5\times 3}. \tag{28}$$

§ 10. Applications of definite integrals

101. Calculation of area. We saw in [87] that the area bounded by a given curve $y = f(x)$, by the axis OX and two ordinates $x = a$ and $x = b$ is, expressed by the definite integral:

$$\int_a^b f(x) \, \mathrm{d}x.$$

The area thus found, however, does not give us the actual sum of the areas, formed by the given curve with OX, but gives only their

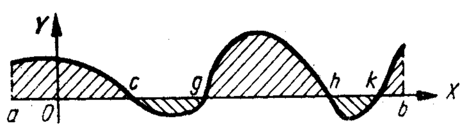

FIG. 127

algebraic sum, in which every area situated below OX takes the (—) sign. To find the sum of these areas in the ordinary sense, we have to find

$$\int_a^b |f(x)| \, \mathrm{d}x.$$

Thus, the sum of the shaded areas in Fig. 127 is equal to

$$\int_a^c f(x) \, \mathrm{d}x - \int_c^g f(x) \, \mathrm{d}x + \int_g^h f(x) \, \mathrm{d}x - \int_h^k f(x) \, \mathrm{d}x + \int_k^b f(x) \, \mathrm{d}x.$$

The area contained between two curves

$$y = f(x), \ y = \varphi(x) \tag{1}$$

and two ordinates

$$x = a, \ x = b \, ,$$

where one curve lies below the other, i.e.

$$f(x) \geqslant \varphi(x)$$

in the interval (a, b), *is given by the definite integral*

$$\int_a^b [f(x) - \varphi(x)] \, dx. \tag{2}$$

We suppose first that both curves lie above OX. It is at once evident from Fig. 128 that the required area S is the difference between the areas bounded by the given curves with OX:

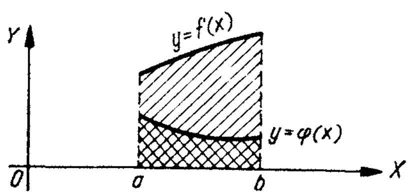

FIG. 128

$$S = \int_a^b f(x) \, dx - \int_a^b \varphi(x) \, dx = $$
$$= \int_a^b [f(x) - \varphi(x)] \, dx,$$

which is what we required to prove. The general case, of any disposition of the two curves relative to OX, now follows by displacing OX downwards, so that both curves appear above OX; this displacement is equivalent to adding the same constant to both functions, with the difference $f(x) - \varphi(x)$ remaining unchanged.

We propose as an exercise the proof that, *if two given curves intercept, i.e. part of one curve is below the other, and part above, the sum of the areas contained between the curves and the ordinates* $x = a$ *and* $x = b$ *is equal to*

$$\int_a^b |f(x) - \varphi(x)| \, dx. \tag{3}$$

Evaluation of a definite integral is often called *quadrature*. This is because finding a definite integral often amounts to finding an area, as shown above.

Examples. 1. *The area bounded by the second degree parabola*

$$y = ax^2 + bx + c \, ,$$

the axis OX, and two ordinates, distant h apart, is equal to

$$\frac{h}{6}(y_1 + y_2 + 4y_0), \tag{4}$$

where y_1 and y_2 denote the outer ordinates of the curve, and y_0 lies equidistant from each of these outer ordinates.

We assume here that the curve lies above OX.

We can take without loss of generality the left-hand outer ordinate as along OY (Fig. 129) in proving (4), since displacement of the entire figure along OX changes neither the magnitude of the area in question, nor the relative disposition of the outer and centre ordinates, nor the size of these ordinates. With this assumption, and with the equation of the parabola in the form:

$$y = ax^2 + bx + c,$$

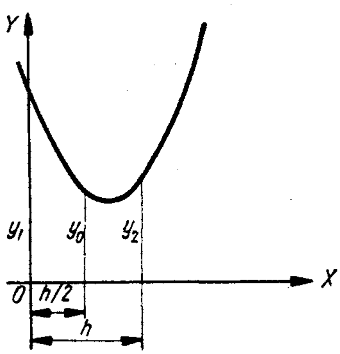

FIG. 129

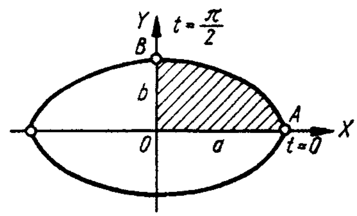

FIG. 130

we write the required area S as the definite integral:

$$S = \int_0^h (ax^2 + bx + c)\,dx = a\frac{x^3}{3} + \frac{bx^2}{2} + cx\Big|_0^h =$$

$$= a\frac{h^3}{3} + b\frac{h^2}{2} + ch = \frac{h}{6}(2ah^2 + 3bh + 6c).$$

We have from our notation:

$$y_0 = ax^2 + bx + c\Big|_{x=\frac{1}{2}h} = \frac{1}{4}ah^2 + \frac{1}{2}bh + c,$$

$$y_1 = ax^2 + bx + c\,|_{x=0} = c, \quad y_2 = ax^2 + bx + c\,|_{x=h} = ah^2 + bh + c,$$

whence it follows:

$$y_1 + y_2 + 4y_0 = 2ah^2 + 3bh + 6c,$$

which proves our example.

2. *Area of an ellipse.* An ellipse having the equation

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

is symmetrical relative to the axes, and hence the required area S is equal to four times the area of the part lying in the first quadrant, i.e.

$$S = 4\int_0^a y\,dx$$

(Fig. 130). Instead of finding y from the equation of the ellipse and substituting the expression obtained in the integrand, we use the parametric form of the ellipse:

$$x = a\cos t\,,\quad y = b\sin t\,, \tag{5}$$

introducing the new variable t in place of x; y is now expressed directly by the second of equations (5). When x varies from 0 to a, t varies from $\pi/2$ to 0, and since all the conditions for the rule of change of variables [99] are fulfilled in this case,

$$S = 4\int_{\pi/2}^{0} b\,\sin t\,d\,(a\,\cos\,t) =$$

$$= -\,4ab\int_{\pi/2}^{0}\sin^2 t\,dt = 4ab\int_0^{\pi/2}\sin^2 t\,dt.$$

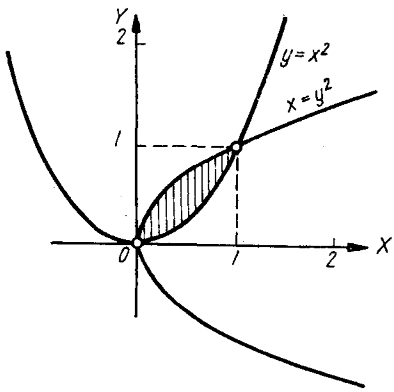

FIG. 131

We have for $k = 1$ from (27) [100]:

$$\int_0^{\pi/2}\sin^2 t\,dt = \frac{1}{2}\cdot\frac{1}{2}\pi = \frac{\pi}{4}\,,$$

so that we finally get:

$$S = \pi ab. \tag{6}$$

With $a = b$, when the ellipse becomes a circle of radius a, we get the well-known expression πa^2 for the area of a circle.

3. *To find the area contained between the two curves*

$$y = x^2\,,\ x = y^2.$$

The given curves (Fig. 131) intercept at two points (0,0), (1,1), their coordinates being obtained by simultaneously solving the equations of the curves. Since we have in (0,1):

$$\sqrt{x} > x^2\,,$$

we find on using (2) that the required area S is given by

$$S = \int_0^1 (\sqrt{x} - x^2)\,dx = \left(\frac{2}{3}x^{3/2} - \frac{x^3}{3}\right)\Bigg|_0^1 = \frac{1}{3}.$$

102. Area of a sector. *The area of the sector bounded by a curve with polar equation*

$$r = f(\theta)\,, \tag{7}$$

and the two radius vectors

$$\theta = \alpha,\ \theta = \beta \tag{8}$$

drawn from the pole at angles α and β to the polar axis, is given by :

$$S = \int_{\alpha}^{\beta} \frac{1}{2} r^2\, d\theta = \frac{1}{2}\int_{\alpha}^{\beta} [f(\theta)]^2\, d\theta. \tag{9}$$

We deduce (9) by dividing the area in question into small elements (Fig. 132), the angle between the radius vectors (8) being divided into n parts. We take the area of one such small sector, bounded by radius vectors θ and $\theta + \Delta\theta$. Let ΔS denote this area, and m and M the least and greatest values respectively of $r = f(\theta)$ in the interval $(\theta, \theta + \Delta\theta)$; then we see that ΔS lies between the areas of two circular sectors with the same angle $\Delta\theta$ but with radii m and M, i.e.

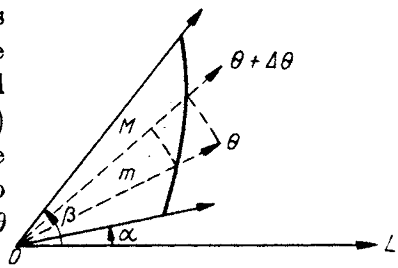

FIG. 132

$$\frac{1}{2} m^2 \Delta\theta \leqslant \Delta S \leqslant \frac{1}{2} M^2 \Delta\theta\,,$$

and hence, denoting some number between m and M by P, we can write:

$$\Delta S = \frac{1}{2} P^2 \Delta\theta.$$

Since the continuous function $f(\theta)$ takes all values between m and M in the interval $(\theta, \theta + \Delta\theta)$, there must exist θ' in this interval such that

$$f(\theta') = P\,,$$

so that now

$$\Delta S = \frac{1}{2} [f(\theta')]^2 \Delta\theta. \tag{10}$$

If we now increase the number of elementary sectors ΔS, so that the greatest of the $\Delta\theta$ tends to zero, and if we recall what was said in [87], we get in the limit:

$$S = \lim \sum \Delta S = \lim \sum \frac{1}{2} [f(\theta')]^2 \Delta\theta = \int_{\alpha}^{\beta} \frac{1}{2} [f(\theta)]^2\, d\theta = \int_{\alpha}^{\beta} \frac{1}{2} r^2\, d\theta\,,$$

which it was required to prove.

We remark that the basic idea in the above proof of (9) lies in replacing the area ΔS by the area of a circular sector of the same angle $\Delta\theta$ and radius $f(\theta')$. If instead of the *exact* expression (10) we use the approximation:

$$\Delta S = \frac{1}{2} r^2 \Delta\theta ,$$

where $r = f(\theta'')$ and θ'' is any value from the interval $(\theta, \theta + \Delta\theta)$, we get in the limit the same result for the area of a sector:

$$\lim \sum \frac{1}{2} [f(\theta'')]^2 \Delta\theta = \int_{\alpha}^{\beta} \frac{1}{2} r^2 \, d\theta . \qquad (11)$$

This leads us to a simple geometrical interpretation of the integral expression in (11): $\frac{1}{2} r^2 \, d\theta$ is an approximate expression for the

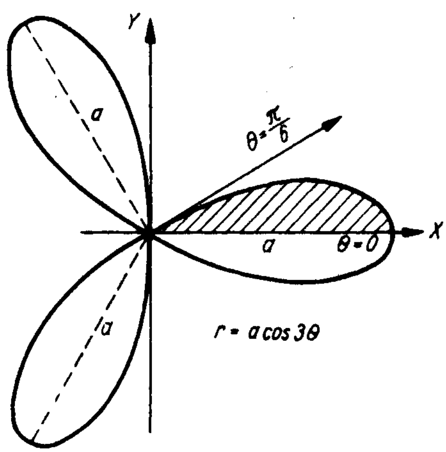

FIG. 133

area of an elementary sector of angle $d\theta$, and hence is simply referred to as an *elementary area in polar coordinates*.

Example. To find the area bounded by the closed curve

$$r = a \cos 3\theta \qquad (a > 0).$$

This curve, called a trifolium, is shown in Fig. 133 and is plotted without difficulty. The total area bounded by it is six times the area of the shaded part, corresponding to a variation of θ from 0 to $\pi/6$; so that we have from (9):

$$S = 6 \int_0^{\pi/6} \frac{1}{2} a^2 \cos^2 3\theta \, d\theta = a^2 \int_0^{\pi/6} \cos^2 3\theta \, d(3\theta) = a^2 \int_0^{\pi/2} \cos^2 t \, dt = \frac{\pi a^2}{4}.$$

103. Length of arc. Let AB be the arc of a certain curve. We draw a series of successive chords (Fig. 134) and increase their number in such a way that the greatest tends to zero. If the total length of the series of chords now tends to a definite limit, independently of how the chords are drawn, the arc is said to be *rectifiable*, and the limit in question is called the *length of this arc*. This definition of length is also suitable for a closed curve.

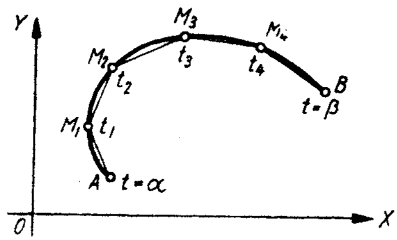

FIG. 134

Let the curve be given explicitly as $y = f(x)$, and let A and B correspond to $x = a$ and $x = b$ respectively $(a < b)$; let $f(x)$ also have a continuous derivative in the interval $a \leqslant x \leqslant b$, to which the arc AB corresponds. We show that the arc AB can be rectified in these conditions, and that its length is expressed by a definite integral.

Let the chords be $AM_1, M_1M_2, \ldots, M_{n-1}B$, and let the coordinates of their ends be

$$a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b,$$

and $y_i = f(x_i)$. Using the formula of analytic geometry, we have for the total length of the chords:

$$p = \sum_{i=1}^{n} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} =$$

$$= \sum_{i=1}^{n} \sqrt{(x_i - x_{i-1})^2 + (f(x_i) - f(x_{i-1}))^2}.$$

Using the formula for finite increments:

$$f(x_i) - f(x_{i-1}) = f'(\xi_i)(x_i - x_{i-1}) \qquad (x_{i-1} < \xi_i < x_i),$$

we get for the length of a single chord

$$\sqrt{1 + f'^2(\xi_1)}\,(x_i - x_{i-1}),$$

from which we see that the requirement that the greatest chord tends to zero is equivalent to the requirement that the greatest of the $(x_i - x_{i-1})$ tends to zero. We now have

$$p = \sum_{i=1}^{n} \sqrt{1 + f'^2(\xi_i)}\,(x_i - x_{i-1}),$$

which in fact has a limit equal to the integral

$$\int_a^b \sqrt{1 + f'^2(x)}\, dx .$$

Thus, the length l of arc AB is given by:

$$l = \int_a^b \sqrt{1 + f'^2(x)}\, dx . \tag{12}$$

Let $x' < x''$ be any two values of x in the interval (a, b), and let M', M'' be the corresponding points of arc AB. Using the mean value theorem, we get the following expression for the length l' of arc $M'M''$:

$$l' = \int_{x'}^{x''} \sqrt{1 + f'^2(x)}\, dx = \sqrt{1 + f'^2(\xi_1)}\,(x'' - x')\ (x' < \xi_1 < x'').$$

Using the formula for finite differences, we get for the length of chord $M'M''$:

$$M'M'' = \sqrt{(x'' - x')^2 + [f(x'') - f(x')]^2} =$$

$$= \sqrt{1 + f'^2(\xi_2)}\,(x'' - x') \qquad (x' < \xi_2 < x'').$$

Hence it follows:

$$\frac{M'M''}{l'} = \frac{\sqrt{1 + f'^2(\xi_2)}}{\sqrt{1 + f'^2(\xi_1)}} .$$

If M' and M'' tend to M with abscissa x, x' and x'' tend to x, and also ξ_1 and $\xi_2 \to x$, and we get from the above formula:

$$\frac{M'M''}{l'} \to 1 .$$

We made use of this in [70].

We now suppose that the curve is given parametrically,

$$x = \varphi(t),\ y = \psi(t) ,$$

where A and B correspond to $t = \alpha$ and $t = \beta$ $(\alpha < \beta)$. We assume that points of the curve AB correspond to t in the interval $\alpha \leqslant t \leqslant \beta$, with different points for different t, and with the curve neither closed nor intercepting itself (Fig. 134). We assume further that there exist continuous derivatives $\varphi'(t)$ and $\psi'(t)$ in the interval $\alpha \leqslant t \leqslant \beta$.

Let chords be drawn as above, $AM_1, M_1M_2, \ldots, M_{n-1}B$, with their ends corresponding to $t_0 = \alpha < t_1 < t_2 < \ldots < t_{n-1} < t_n = \beta$.

We have for the total length of the chords:

$$p = \sum_{i=1}^{n} \sqrt{(\varphi(t_i) - \varphi(t_{i-1}))^2 + (\psi(t_i) - \psi(t_{i-1}))^2}$$

or, using the formula for finite increments,

$$p = \sum_{i=1}^{n} \sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i')}\,(t_i - t_{i-1})\ (t_{i-1} < \tau_i < t_i) \text{ and } (t_{i-1} < \tau_i' < t_i) \quad (13)$$

It can be shown that the requirement that the greatest of the chords tends to zero is equivalent to the requirement that the greatest of $(t_i - t_{i-1})$ tends to zero. This can be proved without assuming the existence of derivatives $\varphi'(t)$ and $\psi'(t)$.

Expression (13) differs from the sum giving in the limit the integral

$$\int_{\alpha}^{\beta} \sqrt{\varphi'^2(t) + \psi'^2(t)}\,dt\,, \quad (14)$$

since the arguments τ_i and τ_i' are in general different. We introduce the sum

$$q = \sum_{i=1}^{n} \sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i)}\,(t_i - t_{i-1})\,,$$

which gives integral (14) in the limit. To show that (13) tends to the limit (14), we have to show that the difference

$$p - q = \sum_{i=1}^{n} \left[\sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i')} - \sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i)}\right](t_i - t_{i-1})$$

tends to zero.

We multiply and divide by the sum of the radicals, and get

$$p - q = \sum_{i=1}^{n} \frac{\psi'(\tau_i') + \psi'(\tau_i)}{\sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i')} + \sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i)}}\,(\psi'(\tau_i') - \psi'(\tau_i))\,(t_i - t_{i-1}).$$

Since

$$(\psi'(\tau_i') + \psi'(\tau_i)) \leqslant \sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i')} + \sqrt{\varphi'^2(\tau_i) + \psi'^2(\tau_i)}\,,$$

we have

$$|p - q| \leqslant \sum_{i=1}^{n} |\psi'(\tau'_i) - \psi'(\tau_i)|\,(t_i - t_{i-1}).$$

The numbers τ_i and τ_i' belong to the interval (t_{i-1}, t_i), and we can say, by the uniform continuity of $\psi'(t)$ in the interval $\alpha \leqslant t \leqslant \beta$, that the

greatest of the $|\psi'(\tau_i') - \psi'(\tau_i)|$, which we denote by δ, will tend to zero if the greatest of the $(t_i - t_{i-1})$ tends to zero. But we have from the above formula:

$$|p - q| \leqslant \sum_{i=1}^{n} \delta (t_i - t_{i-1}) = \delta \sum_{i=1}^{n} (t_i - t_{i-1}) = \delta (\beta - \alpha),$$

whence it is clear that $p - q \to 0$. Thus the total length of the chords given by (13), tends to integral (14), i.e.

$$l = \int_{\alpha}^{\beta} \sqrt{\varphi'^2 (t) + \psi'^2 (t)}\, \mathrm{d}t. \tag{15}$$

This formula for the length l remains valid in the case of a closed curve. This can be shown, for instance, simply by dividing the closed curve into two open curves, applying (16) to each, and adding the values of l obtained. Similarly, if a curve L consists of a finite number of curves L_k, each with a parametric form that satisfies the conditions given above, the length L_k of each can be found from (15), and the total length L obtained by summation.

We take t variable in the interval (α, β), with the corresponding point M varying over the arc AB. The length of arc AM will be a function of t, given by

$$s(t) = \int_{\alpha}^{t} \sqrt{\varphi'^2 (t) + \psi'^2 (t)}\, \mathrm{d}t. \tag{16}$$

We get by the rule for differentiation of an integral with respect to its upper limit,

$$\frac{\mathrm{d}s}{\mathrm{d}t} = \sqrt{\varphi'^2 (t) + \psi'^2 (t)}, \tag{17}$$

i.e.

$$\mathrm{d}s = \sqrt{\varphi'^2 (t) + \psi'^2 (t)}\, \mathrm{d}t,$$

whence, noting that

$$\varphi'(t) = \frac{\mathrm{d}x}{\mathrm{d}t}, \quad \psi'(t) = \frac{\mathrm{d}y}{\mathrm{d}t},$$

we get the formula for the differential of an arc [70]:

$$\mathrm{d}s = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2},$$

whilst (15) can be written without indicating the variable of integration, in the form:

$$l = \int_{(A)}^{(B)} \mathrm{d}s = \int_{(A)}^{(B)} \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2}.$$

Limits (A) and (B) indicate the initial and final points of the arc.

If $\varphi'^2(t) + \psi'^2(t) > 0$ for all t in (α, β), (17) gives us the derivative of the parameter t with respect to s:

$$\frac{dt}{ds} = \frac{1}{\sqrt{\varphi'^2(t) + \psi'^2(t)}}.$$

The existence of continuous derivatives $\varphi'(t)$ and $\psi'(t)$, together with the condition $\varphi'^2(t) + \psi'^2(t) > 0$, ensures a tangent varying continuously along AB.

If the curve is given in polar coordinates by the equation

$$r = f(\theta),$$

we can introduce rectangular coordinates x and y, related to r and θ by

$$x = r\cos\theta, \quad y = r\sin\theta \tag{18}$$

[82], then take these equations as a parametric form of the curve with the parameter θ.

We then have:

$$dx = \cos\theta\, dr - r\sin\theta\, d\theta;$$

$$dy = \sin\theta\, dr + r\cos\theta\, d\theta;$$

$$dx^2 + dy^2 = (dr)^2 + r^2(d\theta)^2,$$

whence

$$ds = \sqrt{(dx)^2 + (dy)^2} = \sqrt{(dr)^2 + r^2(d\theta)^2} \tag{19}$$

and if A, B correspond to values α, β respectively of the polar angle θ (Fig. 135), (15) gives us:

$$s = \int_{\alpha}^{\beta} \left(\sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\right) d\theta. \tag{20}$$

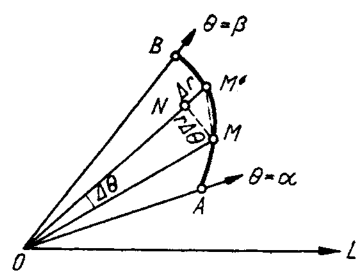

FIG. 135

Expression (19) for ds, called *the differential of an arc in polar coordinates*, can be got directly from the figure, by replacing the infinitely small arc MM' by its chord, then taking the chord as the hypotenuse of the right-angled triangle MNM', the adjacent sides of which, $\overline{MN}$ and $\overline{NM'}$, are respectively approximately equal to $r d\theta$ and dr.

Examples. 1. The length of arc s of the *parabola* $y = x^2$, measured from the vertex (0,0) to a variable point with abscissa x, is given by (12) as the

integral:

$$s = \int_0^x \sqrt{1 + y'^2}\, dx = \int_0^x \sqrt{1 + 4x^2}\, dx = \frac{1}{2}\int_0^{2x} \sqrt{1 + t^2}\, dt \tag{21}$$

(substituting $t = 2x$).

We have by Example 11 [92]:

$$\int \sqrt{1 + t^2}\, dt = \frac{1}{2}\left[t\sqrt{1 + t^2} + \log\left(t + \sqrt{1 + t^2}\right)\right] + C.$$

Substituting this in (21), we easily obtain:

$$s = \frac{1}{4}\left[2x\sqrt{1 + 4x^2} + \log\left(2x + \sqrt{1 + 4x^2}\right)\right].$$

2. The length of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

is equal, by its symmetry with respect to the axes, to four times the length of the part lying in the first quadrant. We use the parametric equations of the ellipse

$$x = a\cos t,\ \ y = b\sin t\,,$$

and note that variation from A to B means variation of the parameter from 0 to $\pi/2$; we then get, by (15), the following expression for the required length l:

$$l = 4\int_0^{\pi/2} \sqrt{a^2\sin^2 t + b^2\cos^2 t}\, dt. \tag{22}$$

This integral cannot be evaluated in an explicit form; a method of approximate evaluation, given later, has to be used.

3. The length of arc of the *logarithmic spiral*

$$r = Ce^{a\theta}$$

[83], cut off by the radius-vectors $\theta = \alpha$, $\theta = \beta$, is given by (20) as the integral:

$$\int_\alpha^\beta \sqrt{r^2 + \left(\frac{dr}{d\theta}\right)^2}\, d\theta = C\sqrt{1 + a^2}\int_\alpha^\beta e^{a\theta}\, d\theta = \frac{C\sqrt{1 + a^2}}{a}\left[e^{a\beta} - e^{a\alpha}\right].$$

4. Let $M(x, y)$ be a point of the catenary considered in [78]. We find the length of arc AM (Fig. 93). Using the expression for $(1 + y'^2)$ in [78], we get:

$$AM = \int_0^x \sqrt{1 + y'^2}\, dx = \int_0^x \frac{y}{a}\, dx = \frac{1}{2}\int_0^x \left(e^{x/a} + e^{-x/a}\right) dx =$$

$$= \frac{a}{2}\left(e^{x/a} - e^{-x/a}\right) = ay'\,,$$

whence

$$a^2 + (\text{arc } AM)^2 = a^2 + a^2 y'^2 = a^2(1 + y'^2) = y^2 ,$$

i.e. the length of arc AM is equal to the adjacent side of a right-angled triangle with hypotenuse equal to the ordinate of M, and the other adjacent side equal to a. This gives us the following rule for finding the length of arc AM:

With the vertex A of the catenary as centre, describe a circle with radius equal to the ordinate of M ; let the circle cut OX in Q ; then OQ is the rectified arc AM (Fig. 93).

We have been guided in the choice of sign in the above formula by the fact that y' is positive for points lying on the right-hand side of the catenary.

5. We take the cycloid of [79] and find the length of arc l of the branch OO' (Fig. 94), and the area S bounded between this branch and OX:

$$l = \int_0^{2\pi} \sqrt{[\varphi'(t)]^2 + [\psi'(t)]^2}\, dt = \int_0^{2\pi} \sqrt{a^2(1-\cos t)^2 + a^2 \sin^2 t}\, dt =$$

$$= a \int_0^{2\pi} \sqrt{2 - 2\cos t}\, dt = a \int_0^{2\pi} \sqrt{4 \sin^2 \frac{1}{2} t}\, dt = 2a \int_0^{2\pi} \sin \frac{1}{2} t\, dt =$$

$$= 2a \left[-2 \cos \frac{1}{2} t \right]_0^{2\pi} = 8a,$$

i.e. *the length of arc of one branch of a cycloid is eight times the diameter of the rolling circle ;*

$$S = \int_0^{2\pi a} y\, dx = \int_0^{2\pi} \psi(t)\, \varphi'(t)\, dt = a^2 \int_0^{2\pi} (1 - \cos t)^2\, dt =$$

$$= a^2 \int_0^{2\pi} (1 - 2\cos t + \cos^2 t)\, dt = 2\pi a^2 - 2a^2 [\sin t]_0^{2\pi} +$$

$$+ a^2 \left[\frac{1}{2} t + \frac{1}{4} \sin 2t \right]_0^{2\pi} = 2\pi a^2 + \pi a^2 = 3\pi a^2.$$

i.e. *the area bounded by one branch of a cycloid and the fixed line along which the circle rolls is equal to three times the area of the rolling circle.*

We had to take the $(+)$ sign with $\sqrt{4 \sin^2 \frac{1}{2} t}$ in finding l, since $\sin \frac{1}{2} t$ is positive with t varying from 0 to 2π.

6. The cardioid of [84] is symmetrical with respect to the polar axis (Fig. 111), and its length l can therefore be found by doubling the length of arc for θ varying from 0 to π:

$$l = 2 \int_0^{\pi} \sqrt{r^2 + r'^2}\, d\theta = 2 \int_0^{\pi} \sqrt{4a^2(1 + \cos\theta)^2 + 4a^2 \sin^2\theta}\, d\theta =$$

$$= 8a \int_0^{\pi} \cos \frac{1}{2} \theta\, d\theta = 8a \left[2 \sin \frac{1}{2} \theta \right]_0^{\pi} = 16a,$$

i.e. *the length of arc of a cardioid is eight times the diameter of the rolling (or fixed) circle.*

104. Calculation of the volumes of solids of given cross-section. The calculation of the volume of a solid also leads to evaluation of a definite integral, when the area of the cross-section, perpendicular to a given direction, is known.

We call the volume of the given solid V (Fig. 136) and we assume that the cross-sectional area of the solid is known in planes perpendi-

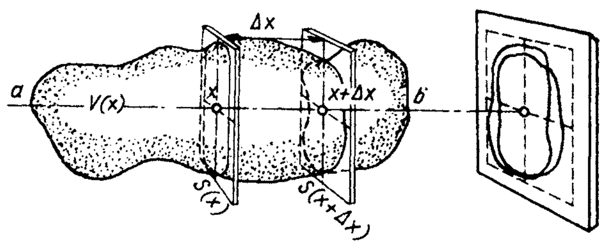

FIG. 136

cular to a given direction, which is taken as OX. Each cross-section is defined by the abscissa x of its point of intersection with OX, so that the cross-sectional area is a function of x, say $S(x)$, which we assume known.

Further, let a, b denote the abscissae of the extreme sections of the solid. We find V by dividing the solid into elements, by means

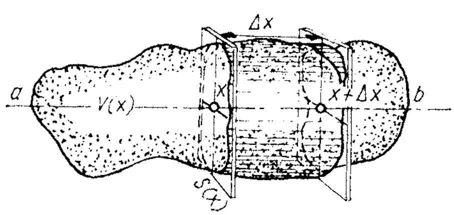

FIG. 137

of a series of cross-sections starting at $x = a$ and ending at $x = b$ we take one such element ΔV between the cross-sections with abscissae x and $x + \Delta x$. We replace ΔV by the volume of a right cylinder of height Δx and base equal to the cross-section of the solid at x (Fig 137). The volume of this cylinder is $S(x)\,\Delta x$, so that the required volume V will be given approximately by

$$\Sigma\, S(x)\, \Delta x\ ,$$

where the summation is over all the elements into which the solid has been divided. On indefinitely increasing the number of elements, and with the greatest of the Δx tending to zero, the above sum becomes in the limit a definite integral, accurately expressing V; this leads to the following proposition:

If the cross-section of a solid is known for all the planes lying perpendicular to a given direction, which we take as the axis OX, the volume V of the solid is given by :

$$V = \int_a^b S(x)\,\mathrm{d}x, \qquad (23)$$

where $S(x)$ denotes the area of the cross-section of abscissa x, and a, b are the abscissae of the extremities of the solid.

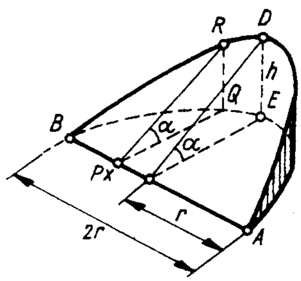

FIG. 138

Example. To find the *volume of a cylindrical ungula,* cut from a right circular cylinder by a plane passing through a diameter of its base (Fig.138). We take the diameter AB as OX, with A as origin; we let r denote the radius of the base of the cylinder, and α the angle between the cutting plane and the base.

A cross-section perpendicular to AB consists of a right-angled triangle PQR, the area of which is:

$$S(x) = \frac{1}{2}\,\overline{PQ}\cdot\overline{QR} = \frac{1}{2}\tan\alpha\cdot\overline{PQ}^2.$$

Further, we know from the properties of circles that $\overline{PQ}$ is the geometric mean of $\overline{AP}$, $\overline{PB}$, i.e.

$$\overline{PQ}^2 = \overline{AP}\cdot\overline{PB} = x(2r - x),$$

so that finally,

$$S(x) = \frac{1}{2}\,x(2r - x)\tan\alpha.$$

Using (23), we get for the required volume V:

$$V = \int_0^{2r} S(x)\,\mathrm{d}x = \frac{1}{2}\tan\alpha\int_0^{2r} x(2r - x)\,\mathrm{d}x = \frac{1}{2}\tan\alpha\left(rx^2 - \frac{x^3}{3}\right)\Big|_0^{2r} =$$

$$= \frac{2}{3}\,r^3\tan\alpha = \frac{2}{3}\,r^2 h,$$

where the "height" of the ungula $h = r\tan\alpha$.

105. Volume of a solid of revolution. When the solid in question is obtained by revolution of a given curve $y = f(x)$ about OX, its cross-section is a circle of radius y (Fig. 139), and hence

$$S(x) = \pi y^2,$$

$$V(x) = \int_a^b \pi y^2 \,\mathrm{d}x,$$

i.e. *the volume of the solid, obtained by revolution about* OX *of the part of the curve*

$$y = f(x)$$

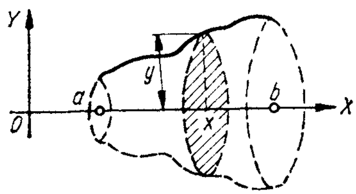

FIG. 139

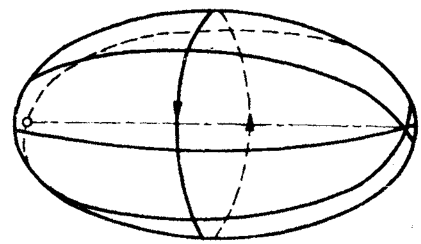

FIG. 140

included between the ordinates $x = a$ *and* $x = b$, *is given by :*

$$V = \int_a^b \pi y^2 \,\mathrm{d}x \tag{24}$$

Example. Volume of the ellipsoid of revolution. We get by revolution of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1$$

about its major axis the *prolate ellipsoid of revolution*. (See Fig. 140). The extreme abscissae x are $(-a)$ and $(+a)$ in this case, so that (24) gives:

$$V_{\text{pro}} = \pi \int_{-a}^{+a} y^2 \,\mathrm{d}x = \pi \int_{-a}^{+a} b^2 \left(1 - \frac{x^2}{a^2}\right) \mathrm{d}x =$$

$$= \pi b^2 \left(x - \frac{x^3}{3a^2}\right) \Bigg|_{-a}^{+a} = \frac{4}{3}\pi a b^2. \tag{25}$$

The volume of the *oblate ellipsoid of revolution*, obtained by revolution of the ellipse about its minor axis, is found in the same way. It is only necessary

to interchange x, y and a, b, which gives:

$$V_{\text{ob}} = \pi \int_{-b}^{+b} x^2\,\mathrm{d}y = \pi \int_{-b}^{+b} a^2\left(1 - \frac{y^2}{b^2}\right)\mathrm{d}y = \frac{4}{3}\pi b a^2 \tag{26}$$

When $a = b$, both ellipsoids reduce to a sphere of radius a, with volume $\frac{4}{3}\pi a^3$.

106. Surface area of a solid of revolution. *When a solid is obtained by the revolution of a curve, given in the XOY plane, about the axis OX, its surface area is defined as the limit of the surface area of a second solid, the second solid being obtained by revolution of a successive series of chords,*

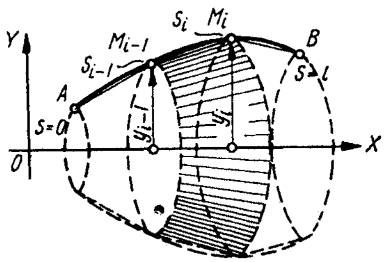

FIG. 141

inscribed in the original curve, also about OX, the limit being taken when the number of chords increases indefinitely, and the greatest of them tends to zero (Fig. 141).

If the part of the rotating curve lies between A and B, the surface area F is given for the solid of revolution by :

$$F = \int_{(A)}^{(B)} 2\pi y\,\mathrm{d}s, \tag{27}$$

where ds *is the differential of the arc of the given curve*, i.e.

$$\mathrm{d}s = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2}.$$

The curve may be given in any convenient form in the above formula, either explicitly or parametrically; symbols (A), (B) indicate that integration is between the limits of the independent variable corresponding to the given points A, B of the curve.

We shall assume that the equation of the curve is given parametrically, with the length of arc s as parameter, measured from the point A;

let l denote the total length of curve AB. The curve is of course assumed to be rectifiable. We have: $x = \varphi(s)$, $y = \psi(s)$. We divide the interval $(0, l)$ of variation of s into sub-intervals in the usual way, with

$$0 = s_0 < s_1 < s_2 < \ldots < s_{n-1} < s_n = l.$$

Let $s = s_i$ correspond to the point M_i of the curve, with M_0 evidently coinciding with A, and M_n with B. Let q_i denote the length of the chord $M_{i-1} M_i$, and Δs_i the length of the arc $M_{i-1} M_i$, and let $y_i = \psi(s_i)$. Using the formula for the lateral surface of the frustrum of a cone, we have the following result for the surface area obtained by revolution of the series of chords $AM_1 M_2 \ldots M_{n-1} B$:

$$Q = 2\pi \sum_{i=1}^{n} \frac{1}{2} (y_{i-1} + y_i) q_i$$

or

$$Q = 2\pi \sum_{i=1}^{n} y_{i-1} q_i + \pi \sum_{i=1}^{n} (y_i - y_{i-1}) q_i.$$

Let δ be the greatest of the $|y_i - y_{i-1}|$. By the uniform continuity of $\psi(s)$ in $0 \leqslant s \leqslant l$, δ tends to zero if the greatest value of $(s_i - s_{i-1})$ tends to zero. But we have:

$$\left| \sum_{i=1}^{n} (y_i - y_{i-1}) q_i \right| \leqslant \delta \sum_{i=1}^{n} q_i \leqslant \delta l,$$

whence it follows that the second term in the expression for Q tends to zero. We consider the first term by rewriting it as:

$$2\pi \sum_{i=1}^{n} y_{i-1} q_i = 2\pi \sum_{i=1}^{n} y_{i-1} \Delta s_i - 2\pi \sum_{i=1}^{n} y_{i-1} (\Delta s_i - q_i).$$

We show that the second term on the right-hand side tends to zero, by remarking that $\psi(s)$, continuous in $(0, l)$, is bounded, so that there exists a positive m, such that $|y_{i-1}| \leqslant m$ for all i. Hence:

$$\left| \sum_{i=1}^{n} y_{i-1} (\Delta s_i - q_i) \right| \leqslant \sum_{i=1}^{n} m (\Delta s_i - q_i) = m \left(l - \sum_{i=1}^{n} q_i \right).$$

But if the greatest of the $(s_i - s_{i-1})$ values tends to zero, the greatest of the chords q_i also tends to zero, and the total length of the chords tends to the length of the arc:

$$\sum_{i=1}^{n} q_i \to l,$$

whence

$$2\pi \sum_{i=1}^{n} y_{i-1}(\Delta s_i - q_i) \to 0.$$

The only remaining term to be investigated in the expression for Q is thus:

$$2\pi \sum_{i=1}^{n} y_{i-1}\Delta s_i = 2\pi \sum_{i=1}^{n} \psi(s_{i-1})(s_i - s_{i-1}).$$

But the limit of this sum gives us integral (27). Our formula has thus been proved. If the curve is given in terms of a parameter t, we have [cf. 103]:

$$F = \int_{\alpha}^{\beta} 2\pi\psi(t)\sqrt{\varphi'^2(t) + \psi'^2(t)}\,\mathrm{d}t \qquad (28_1)$$

and in the case of the explicit $y = f(x)$ equation for AB:

$$F = \int_{a}^{b} 2\pi f(x)\sqrt{1 + f'^2(x)}\,\mathrm{d}x. \qquad (28_2)$$

Example. Surface area of prolate and oblate ellipsoid of revolution.

We take the prolate case first. Using the notation of the example of [105], we have by (28):

$$F_{\text{pro}} = 2\pi \int_{-a}^{a} y\sqrt{1 + y'^2}\,\mathrm{d}x = 2\pi \int_{-a}^{a} \sqrt{y^2 + (yy')^2}\,\mathrm{d}x.$$

We have from the equation of the ellipse:

$$y^2 = b^2\left(1 - \frac{x^2}{a^2}\right), \qquad yy' = -\frac{b^2 x}{a^2},$$

whence

$$(yy')^2 = \frac{b^4 x^2}{a^4},$$

$$F_{\text{pro}} = 2\pi \int_{-a}^{a} \sqrt{b^2 - \frac{b^2 x^2}{a^2} + \frac{b^4 x^2}{a^4}}\,\mathrm{d}x = 2\pi b \int_{-a}^{a} \sqrt{1 - \frac{x^2}{a^2}\left(1 - \frac{b^2}{a^2}\right)}\mathrm{d}x.$$

We introduce here the eccentricity of the ellipse

$$\varepsilon^2 = \frac{a^2 - b^2}{a^2},$$

and get (cf. Example [99]):

$$F_{\text{pro}} = 2\pi b \int_{-a}^{a} \sqrt{1 - \frac{\varepsilon^2 x^2}{a^2}}\, \mathrm{d}x = 4\pi b \int_0^a \sqrt{1 - \frac{\varepsilon^2 x^2}{a^2}}\, \mathrm{d}x =$$

$$= \frac{4\pi b a}{\varepsilon} \int_0^a \sqrt{1 - \left(\frac{\varepsilon x}{a}\right)^2}\, \mathrm{d}\left(\frac{\varepsilon x}{a}\right) = \frac{4\pi ab}{\varepsilon} \int_0^\varepsilon \sqrt{1 - t^2}\, \mathrm{d}t.$$

Integrating by parts, we have (cf. Example 11 [92]):

$$\int \sqrt{1 - t^2}\, \mathrm{d}t = t\sqrt{1 - t^2} + \int \frac{t^2}{\sqrt{1 - t^2}} \mathrm{d}t =$$

$$= t\sqrt{1 - t^2} - \int \sqrt{1 - t^2}\, \mathrm{d}t + \int \frac{\mathrm{d}t}{\sqrt{1 - t^2}},$$

whence

$$\int \sqrt{1 - t^2}\, \mathrm{d}t = \frac{1}{2}\left[t\sqrt{1 - t^2} + \arcsin t\right],$$

and finally

$$F_{\text{pro}} = 2\pi ab \left[\sqrt{1 - \varepsilon^2} + \frac{\arcsin \varepsilon}{\varepsilon}\right] \tag{29}$$

This formula is also valid in the limit for $\varepsilon = 0$, i.e. when $b = a$, and the ellipse reduces to a sphere of radius a. The square brackets contain an indeterminate form in this case, which may be evaluated as [65]:

$$\left.\frac{\arcsin \varepsilon}{\varepsilon}\right|_{\varepsilon=0} = \left.\frac{1/\sqrt{1 - \varepsilon^2}}{1}\right|_{\varepsilon=0} = 1.$$

We now turn to the oblate ellipsoid of revolution. We interchange x, y, and a, b, and find:

$$F_{\text{ob}} = 2\pi \int_{-b}^{b} \sqrt{x^2 + (xx')^2}\, \mathrm{d}y,$$

where x is taken as a function of y.

But we have from the equation of the ellipse:

$$x^2 = a^2\left(1 - \frac{y^2}{b^2}\right), \quad xx' = -\frac{a^2 y}{b^2}, \quad (xx')^2 = \frac{a^4 y^2}{b^4},$$

whence

$$F_{\text{ob}} = 2\pi a \int_{-b}^{+b} \sqrt{1 + \frac{y^2}{b^2}\left(\frac{a^2}{b^2} - 1\right)}\, \mathrm{d}y = 4\pi a \int_0^b \sqrt{1 + \frac{y^2 a^2 \varepsilon^2}{b^4}}\, \mathrm{d}y =$$

$$= 4\pi \frac{b^2}{\varepsilon} \int_0^{a\varepsilon/b} \sqrt{1 + t^2}\, \mathrm{d}t = 2\pi \frac{b^2}{\varepsilon}\left[t\sqrt{1 + t^2} + \log\left(t + \sqrt{1 + t^2}\right)\right]\Bigg|_0^{a\varepsilon/b} =$$

$$= 2\pi \frac{b^2}{\varepsilon}\left[\frac{a\varepsilon}{b}\sqrt{1 + \frac{a^2\varepsilon^2}{b^2}} + \log\left(\frac{a\varepsilon}{b} + \sqrt{1 + \frac{a^2\varepsilon^2}{b^2}}\right)\right] =$$

$$= 2\pi \frac{b^2}{\varepsilon}\left[\frac{a\varepsilon}{b}\sqrt{\frac{a^2}{b^2}} + \log\left(\frac{a\varepsilon}{b} + \sqrt{\frac{a^2}{b^2}}\right)\right] = 2\pi a^2 + \frac{2\pi b^2}{\varepsilon} \log \frac{a(1 + \varepsilon)}{b},$$

and finally:

$$F_{\text{ob}} = 2\pi a^2 + \frac{2\pi b^2}{\varepsilon} \log \frac{a(1+\varepsilon)}{b}. \tag{30}$$

107. Determination of centre of gravity. Guldin's theorem. Given a system of n point-masses at

$$M_1(x_1, y_1), M_2(x_2, y_2), \ldots, M_n(x_n, y_n),$$

with respective masses

$$m_1, m_2, \ldots, m_n,$$

the centre of gravity G of the system is defined as the point whose coordinates x_G, y_G satisfy the conditions:

$$Mx_G = \sum_{i=1}^{n} m_i x_i, \quad My_G = \sum_{i=1}^{n} m_i y_i, \tag{31}$$

where M denotes the total mass of the system:

$$M = \sum_{i=1}^{n} m_i.$$

The points of a system may be arranged in any desired manner when finding its centre of gravity G; the aim is to group the points in such a way that the points of any one group may be replaced by a single point at the centre of gravity of the group, the mass at the point being the total mass of the group.

We shall not dwell on the proof of this general principle, since it presents no difficulty and can easily be verified for the simplest particular cases of systems of three, four points etc.

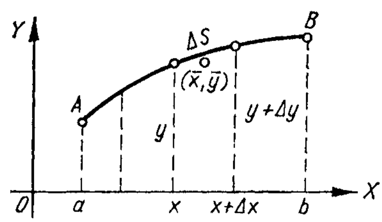

FIG. 142

We shall be concerned in future, not with systems of points, but with the continuous distribution of mass over a plane figure (domain) or on a line.

For simplicity, we limit our consideration to homogeneous solids, the density of which we take as unity, so that the mass is equal to the length in the case of linear distributions, and to the area for plane distributions.

Suppose first that it is required to find the centre of gravity of an arc AB of a curve (Fig. 142), the length of which is s. Following the

above general principle, we divide arc AB into n small elements Δs. The centre of gravity of the total system can be found by replacing each of these elements with its centre of gravity at which is concentrated the total mass of the element $\Delta m = \Delta s$. †

We take an element Δs, and let the coordinates of its ends be (x, y), $(x + \Delta x, y + \Delta y)$; the coordinates of its centre of gravity are denoted by $(\bar{x}, \bar{y})$. We can suppose that the point $(\bar{x}, \bar{y})$ differs by as little as may be desired from the point (x, y) on sufficiently diminishing s.

We have by (31), as in [104]:

$$Mx_G = sx_G = \sum \bar{x}\Delta m = \sum \bar{x}\Delta s = \lim \sum x\Delta s = \int_{(A)}^{(B)} x\mathrm{d}s\,, \tag{32}$$

$$My_G = sy_G = \sum \bar{y}\Delta m = \sum \bar{y}\Delta s = \lim \sum y\Delta s = \int_{(A)}^{(B)} y\mathrm{d}s\,, \tag{33}$$

whence, on obtaining s from the formula:

$$s = \int_{(A)}^{(B)} \mathrm{d}s = \int_{(A)}^{(B)} \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2}\,,$$

we find the coordinates of the centre of gravity G.

An important theorem follows from (32) and (33):

Guldin's first theorem. *The surface area of a solid, obtained by revolution of the arc of a given plane curve about some non-intersecting axis in the same plane, is equal to the product of the length of arc and the path described on revolution by the centre of gravity of the arc.*

Taking the axis of revolution as OX, we have for the surface area F of the solid obtained by revolution of arc AB (using (27) [106]):

$$F = 2\pi \int_{(A)}^{(B)} y\mathrm{d}s = 2\pi y_G \cdot s$$

[by (33)], which it was required to prove.

We now take a plane domain S, its area being also denoted by S. We suppose for simplicity that the domain (Fig. 143) is bounded by two curves, the ordinates of which are denoted by

$$y_1 = f_1(x), \; y_2 = f_2(x)\,.$$

† The centre of gravity of an element does not in general lie on the curve, though it approaches closer to the curve, the smaller the element. This is indicated schematically in Fig. 142.

In accordance with the general principle mentioned at the beginning of this article, we divide the figure into n vertical strips ΔS by lines parallel to OY. We can find the coordinates of the centre of gravity G by replacing each strip by its centre of gravity, associated with its mass $\Delta m = \Delta S$. We take the strip bounded by M_1M_2 and $M_1'M_2'$, with abscissae x and $x + \Delta x$, and let its centre of gravity be $(\bar{x}, \bar{y})$.

On sufficiently reducing the breadth Δx of the strip, $(\bar{x}, \bar{y})$ will differ as little as desired from the centre P of M_1M_2, so that we can write the approximate equations:

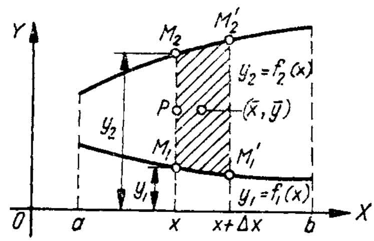

FIG. 143

$$\bar{x} \sim x, \quad \bar{y} \sim \frac{y_1 + y_2}{2}.$$

Further, the mass Δm of the strip is equal to its area ΔS and so can be taken as approximately equal to the area of the rectangle of base Δx and height differing by as little as desired from $\overline{M_1M_2} = y_2 - y_1$, i.e.

$$\Delta m \sim (y_2 - y_1)\, \Delta x.$$

Using (31), we can write:

$$Mx_G = Sx_G = \sum \bar{x}\Delta m = \lim \sum [x(y_2 - y_1)]\, \Delta x = \int_a^b x(y_2 - y_1)\, \mathrm{d}x, \quad (34)$$

$$My_G = Sy_G = \sum \bar{y}\Delta m = \lim \sum \left(\frac{y_2 + y_1}{2}\right)(y_2 - y_1)\, \Delta x =$$

$$= \lim \sum \left[\frac{1}{2}(y_2^2 - y_1^2)\right] \Delta x = \int_a^b \frac{1}{2}(y_2^2 - y_1^2)\, \mathrm{d}x. \quad (35)$$

GULDIN'S SECOND THEOREM. This follows from (35):

When a plane figure revolves about an axis in its plane which does not cut the figure, the volume of the solid obtained is equal to the product of the area of the figure and the length of path resulting from the revolution of the centre of gravity of the figure.

We take the axis of revolution as OX, and notice that the volume V of the solid of revolution in question is equal to the difference

between the volumes of the solids obtained by revolution of the curves y_2 and y_1, so that, in accordance with (24) [105]:

$$V = \pi\int_a^b y_2^2\,\mathrm{d}x - \pi\int_a^b y_1^2\,\mathrm{d}x = \pi\int_a^b (y_2^2 - y_1^2)\,\mathrm{d}x = 2\pi y_G\cdot S\,,$$

by (35), which is what we required to prove.

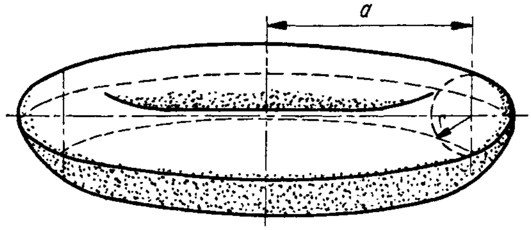

FIG. 144

The above two theorems of Guldin are extremely useful, either for finding the surface area or volume of a solid of revolution, when the centre of gravity of the revolving figure is known, or conversely for finding the centre of gravity of a figure when the volume or surface area of the solid obtained by its revolution is known.

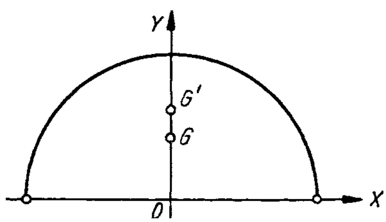

FIG. 145

Examples. **1.** To find the volume V of the anchor ring *(torus)* obtained by revolution of a circle of radius r (Fig. 144) about an axis in its plane, distant a from the centre (with $r < a$, i.e. the axis of revolution does not cut the circle.)

The centre of gravity of the revolving circle evidently lies at its centre, so that the path described by revolution of the centre of gravity is equal to $2\pi a$. The area of the revolving figure is πa^2, and hence we have by Guldin's second theorem:

$$V = \pi r^2\cdot 2\pi a = 2\pi^2 a r^2. \tag{36}$$

2. To find the *surface area* F of the anchor ring of Example 1.

The length of the revolving circle is $2\pi r$; the centre of gravity coincides with the centre of the circle, as before, so that we have by Guldin's first theorem:

$$F = 2\pi r\cdot 2\pi a = 4\pi^2 a r. \tag{37}$$

3. To find the centre of gravity G of a *semi-circular disc* of radius a. We take the base of the semi-circle as axis OX, with OY upwards through the centre and perpendicular to OX (Fig. 145); it is clear from the symmetry of

the figure with respect to OY that G lies on OY. It only remains to find y_G. We apply Guldin's second theorem. The solid obtained by revolution of the semi-circle about OX is a sphere of radius a, and its volume is $\frac{4}{3}\pi a^3$. The area S of the revolving figure is $\frac{1}{2}\pi a^2$, and hence:

$$\frac{4}{3}\pi a^3 = \frac{1}{2}\pi a^2 \times 2\pi y_G, \quad y_G = \frac{4}{3}\frac{a}{\pi}.$$

4. To find the centre of gravity G' of a *semi-circular arc* of radius a.

We take the same axes as in the previous example, and see that G' also lies on OY, so that we have to find $y_{G'}$. We apply Guldin's first theorem, and note that the surface area F of the solid of revolution is $4\pi a^2$ in this case, whilst $s = \pi a$, so that:

$$4\pi a^2 = \pi a \cdot 2\pi y_{G'}, \quad y_{G'} = 2\frac{a}{\pi}.$$

As might be expected, the centre of gravity of the semi-circular arc lies closer to it than the centre of gravity of the semi-circular disc bounded by it.

108. Approximate evaluation of definite integrals; rectangle and trapezoid formulae. It is not always possible to evaluate definite integrals with the aid of the primitive on the basis of (15) [96], since the primitive cannot always be found in practice, even though it exists, given the continuity of the integrand; furthermore, even when it can be found, its form is often too complicated for handy computation. This is the reason for the importance of methods of approximate evaluation of definite integrals.

Most methods are based on the interpretation of a definite integral as an area or as the limit of a sum:

$$\int_a^b f(x)\,dx = \lim \sum_{i=1}^{n} f(\xi_1)(x_i - x_{i-1}). \tag{38}$$

From now on, we agree always to divide the interval (a, b) into n *equal* parts; we let h denote the length of each part, so that

$$h = \frac{b-a}{n}, \quad x_i = a + ih \quad (x_0 = a; \quad x_n = a + nh = b).$$

We also let y_i denote the value of the integrand $y = f(x)$ for $x = x_i$ $(i = 0, 1, \ldots, n)$:

$$y_i = f(x_i) = f(a + ih). \tag{39}$$

We take these values as known; they can be found by direct computation, if $f(x)$ is given analytically, or from the graph, if it is given graphically.

Substituting

$$\xi_i = x_{i-1} \quad \text{or} \quad x_i$$

in the sum on the right-hand side of (38), we get two approximate *rectangle formulae :*

$$\int_a^b f(x)\,\mathrm{d}x \sim \frac{b-a}{n}[y_0 + y_1 + \cdots + y_{n-1}], \tag{40}$$

$$\int_a^b f(x)\,\mathrm{d}x \sim \frac{b-a}{n}[y_1 + y_2 + \cdots + y_n], \tag{41}$$

where the sign ($\sim$) denotes approximate equality.

The greater n, i.e. the smaller h, the more accurate these formulae; in the limit, with $n \to \infty$ and $h \to 0$, they give the exact value of the definite integral.

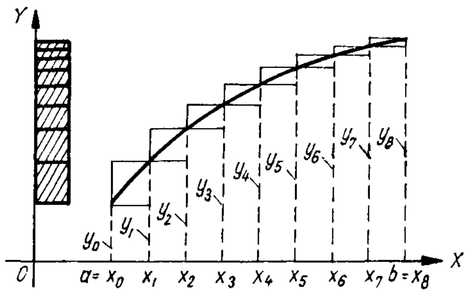

FIG. 146

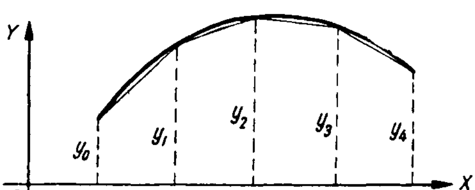

FIG. 147

Thus, *as the number of ordinates increases, the errors of* (40) and (41) *tend to zero.* The upper limit of the error, for a given number of ordinates, is particularly easy to determine for the case of $f(x)$ monotonic in (a, b) (Fig. 146). It is immediately evident from the figure that here, the *error* of each of formulae (40) and (41) *does not exceed* the sum of the areas of the shaded rectangles, i.e. *does not exceed* the area of the rectangle with the same base $(b - a)/n = h$ and with a height equal to the sum of the heights of the shaded rectangles $y_n - y_0$, i.e. the amount:

$$\frac{b-a}{n}(y_n - y_0)\,. \tag{42}$$

The rectangle formula replaces the accurate expression for the area under the curve $y = f(x)$ with an approximation for it, namely the area under the stepline, consisting of horizontal and vertical sections, bounding the rectangles.

We get different approximations by taking other lines, differing by a sufficiently small amount from the given curve, in place of the step-line; the closer the auxiliary line comes to the curve $y = f(x)$, the smaller the error in taking the area under the auxiliary line in place of the area under the curve.

For instance, if we replace the given curve with a series of successive chords with their ends at $x = x_i$ (Fig. 147), so that the required area is replaced

by the sum of the areas of the corresponding trapezoids, we get the approximate *trapezoid formula* :

$$\int_a^b f(x)\,dx \sim h\left[\frac{y_0 - y_1}{2} + \frac{y_1 + y_2}{2} + \ldots + \frac{y_{n-1} + y_n}{2}\right] =$$

$$= \frac{b-a}{2n}\,[y_0 + 2y_1 + 2y_2 + \ldots + 2y_{n-1} + y_n]. \tag{43}$$

109. Tangent formula, and Poncelet's formula. We now double the number of divisions, by halving each sub-interval. We thus get $2n$ divisions (Fig. 148):

$$x_0,\quad x^{1/2} = a + \frac{1}{2}h,$$

$$x_1 = a + h,\quad \ldots,\quad x_i = a + ih,$$

$$x_{i+1/2} = a + \left(i + \frac{1}{2}\right)h,\quad \ldots,\quad x_n = b,$$

with corresponding ordinates:

$y_0, y_{1/2}, y_1, \ldots, y_i, y_{i+1/2}, \ldots, y_n$ (we shall refer to $y_0, y_1, \ldots, y_n$ as *even* ordinates, and to $y_{1/2}$, $y_{3/2}, \ldots, y_{n-1/2}$ as *odd*).

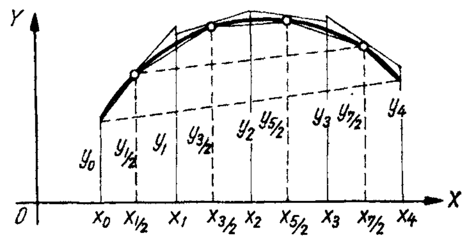

FIG. 148

We draw the tangent at the end of each odd ordinate to cut the two neighbouring even ordinates, and we replace the given area with the sum of the areas of the trapezoids thus constructed. The approximation obtained is referred to as the *tangent formula* :

$$\int_a^b f(x)\,dx \sim \frac{b-a}{n}\,[y_{1/2} + y_{3/2} + \ldots + y_{n-1/2}] = \sigma_1. \tag{44}$$

We consider, together with the above *circumscribed* trapezoids, the *inscribed* trapezoids, obtained by joining the ends of neighbouring odd ordinates with chords; we add to these the two extreme trapezoids, formed by the chords that join the ends of ordinates y_0 and $y_{1/2}$, $y_{n-1/2}$ and y_n. We denote the sum of the areas of these trapezoids by

$$\sigma_2 = \frac{b-a}{2n}\left[\frac{y_0 + y_n}{2} - \frac{y_{1/2} + y_{n-1/2}}{2} + 2y_{1/2} + 2y_{3/2} + \ldots + 2y_{n-1/2}\right].$$

If the curve $y = f(x)$ has no point of inflexion in the interval (a, b), i.e. *is only convex or only concave*, the area S of the curve lies between areas σ_1 and σ_2, so that their arithmetic mean $\frac{1}{2}(\sigma_1 + \sigma_2)$ is naturally taken as an approximation for S. This gives us *Poncelet's formula* :

$$\int_a^b f(x)\,dx \sim \frac{b-a}{2n}\left[\frac{y_0 + y_n}{4} - \frac{y_{1/2} + y_{n-1/2}}{4} + 2y_{1/2} + \right.$$

$$\left. + 2y_{3/2} + \ldots + 2y_{n-1/2}\right]. \tag{45}$$

It is easily seen that, with the assumptions made about the form of the curve, this formula gives an error with an absolute value not exceeding

$$\frac{\sigma_1 - \sigma_2}{2} = \left(\frac{y_{1/2} + y_{n-1/2}}{2} - \frac{y_0 + y_n}{2}\right) \cdot \frac{b-a}{4n}, \tag{46}$$

where it can easily be shown, by considering the bisector of a trapezoid, that the expression in brackets is equal to the segment of the central ordinate cut off by the chords joining the ends of the extreme odd ordinates, and the ends of the extreme even ordinates.

110. Simpson's formula. Still using the previous division into an even number of parts, we replace the given curve by a series of arcs of second degree parabolas, drawn through the ends of each group of three ordinates:

$$y_0, y_{1/2}, y_1; \quad y_1, y_{3/2}, y_2; \ldots; \quad y_{n-1}, y_{n-1/2}, y_n.$$

We use (4) [101] to find the area of each of the curvilinear figures thus obtained, and arrive at Simpson's approximate formula:

$$\begin{aligned}\int_a^b f(x)\,dx \sim \frac{b-a}{6n}[y_0 + 4y_{1/2} + 2y_1 + 4y_{3/2} + \\ + 2y_2 + \ldots + 2y_{n-2} + 4y_{n-1/2} + y_n].\end{aligned} \tag{47}$$

We shall not dwell on deducing the error of this formula, or that of the trapezoid formula. It may be remarked generally, that expressing the error in terms of a definite formula is of theoretical rather than practical interest, since it usually gives too crude a limit.

We note in regard to the above construction that, with suitable choice of a, b, and c, a parabola $y = ax^2 + bx + c$ can always be made to pass through three points of a plane with different abscissae.

Smoothness of the curve is vital for the accuracy of practical results, and the computation should be more precise in the vicinity of points where the curve changes shape fairly rapidly than over intervals of more gradual change. It is always useful to draw at least a rough sketch of the curve before starting the computation.

It is essential to make up a *table* when carrying out numerical computations. The following examples are given to show the arrangement of the tables and in order to compare the accuracy of the various approximate formulae mentioned above:

$$S = \int_0^{\pi/2} \sin x\,dx = 1,$$

$$n = 10, \quad \frac{b-a}{n} = 0.15707963, \qquad \frac{b-a}{2n} = 0.07853981, \qquad \frac{b-a}{6n} = 0.02617994$$

y_1	sin 9°	0.156 4345	$y_{1/2}$	sin 4°.5	0.078 4591
y_2	sin 18°	0.309 0170	$y_{3/2}$	sin 13°.5	0.233 4454
y_3	sin 27°	0.453 9905	$y_{5/2}$	sin 22°.5	0.382 6834
y_4	sin 36°	0.587 7853	$y_{7/2}$	sin 31°.5	0.522 4986
y_5	sin 45°	0.707 1068	$y_{9/2}$	sin 40°.5	0.649 4480
y_6	sin 54°	0.809 0170	$y_{11/2}$	sin 49°.5	0.760 4060
y_7	sin 63°	0.891 0065	$y_{13/2}$	sin 58°.5	0.852 6402
y_8	sin 72°	0.951 0565	$y_{15/2}$	sin 67°.5	0.923 8795
y_9	sin 81°	0.987 6883	$y_{17/2}$	sin 76°.5	0.972 3699
			$y_{19/2}$	sin 85°.5	0.996 9173
Σ_1		5.853 1024	Σ_2		6.372 7474

y_0	sin 0°	0.000 0000
y_{10}	sin 90°	1.000 0000

Low Value Rectangle Formula

$\sum_l$	5.853 1024
y_0	0.000 0000
Σ	5.853 1024

$\log \sum$	0.767 3861
$\log \dfrac{b-a}{n}$	$\bar{1}.196\ 1198$
$\log S$	$\bar{1}.963\ 5059$

$S = 0.919\ 4080$

High Value Rectangle Formula

$\sum_1$	5.853 1024
y_{10}	1.000 0000
Σ	6.853 1024

$\log \sum$	0.835 8873
$\log \dfrac{b-a}{n}$	$\bar{1}.196\ 1198$
$\log S$	0.032 0071

$S \sim 1.076\ 5828$

Tangent Formula

$\log \sum_2$	0.804 3267
$\log \dfrac{b-a}{n}$	$\bar{1}.196\ 1198$
$\log S$	0.000 4465

$S \sim 1.001\ 0290$

Trapezoid Formula

$2\sum_1$	11.706 2048
$y_0 + y_{10}$	1.000 0000
Σ	12.706 2048

$\log \sum$	1.104 0158
$\log \dfrac{b-a}{2n}$	$\bar{2}.895\ 0899$
$\log S$	$\bar{1}.999\ 1057$

$S \sim 0.997\ 9430$

PONCELET'S FORMULA

$2\sum_2$	12.745 4948		$\log \sum$	1.104 7141
$\frac{1}{4}(y_0 + y_{10})$	0.250 0000		$\log \frac{b-a}{2n}$	$\bar{2}.895\,0898$
$-\frac{1}{4}(y_{1/2} + y_{19/2})$	− 0.268 8441			
Σ	12.726 6507		$\log S$	1.999 8039

$$S \sim 0.999\,5487$$

SIMPSON'S FORMULA

$2\sum_1$	11.706 2048		$\log \sum$	1.582 0314
$4\sum_2$	25.490 9896		$\log \frac{b-a}{6a}$	$\bar{2}.417\,9685$
$y_0 + y_{10}$	1.000 0000			
Σ	38.197 1944		$\log S$	$\bar{1}.999\,9999$

$$S \sim 1.000\,0000$$

$$S = \int_0^1 \frac{\log(1+x)}{1+x^2}\,dx = \frac{\pi}{8}\log 2 = 0.272\,198\,2613\ldots, \dagger$$

$$n = 10\,,\ \frac{b-a}{2n} = \frac{1}{20}\,,\ \frac{b-a}{6n} = \frac{1}{60}\,.$$

y_1	0.094 3665	$y_{1/2}$	0.048 6685
y_2	0.175 3092	$y_{3/2}$	0.136 6865
y_3	0.240 7012	$y_{5/2}$	0.210 0175
y_4	0.290 0623	$y_{7/2}$	0.267 3538
y_5	0.324 3721	$y_{9/2}$	0.308 9926
y_6	0.345 5909	$y_{11/2}$	0.336 4722
y_7	0.356 1263	$y_{13/2}$	0.352 0389
y_8	0.358 4065	$y_{15/2}$	0.358 1540
y_9	0.354 6154	$y_{17/2}$	0.357 1470
		$y_{19/2}$	0.351 0273
Σ_1	2.539 5503	Σ_2	2.726 5583

y_0	0.000 0000
y_{10}	0.346 5736

† This formula is derived in Volume II

Poncelet's Formula

$2\sum'_2$	5.453 1166
$\frac{1}{4}(y_9 + y_{10})$	0.086 6434
$-\frac{1}{4}(y_{1/2} + y_{19/2})$	$-$ 0.039 9239
Σ	5.439 8361

$$S = \frac{1}{20}\sum \sim 0.271\,9918$$

Simpson's Formula

$2\sum_1$	5.079 1006
$4\sum_2$	10.906 2332
$y_0 + y_n$	0.346 5736
Σ	16.331 9074

$$S = \frac{1}{60}\sum \sim 0.272\,198\,46$$

$$S = \int_0^1 \frac{dx}{1+x} = \log 2 = 0.693\,147\,18\ldots$$

$$n = 20\,,\ \frac{b-a}{2n} = \frac{1}{40}\,,\ \frac{b-a}{6n} = \frac{1}{120}.$$

y_1	0.952 3810
y_2	0.909 0909
y_3	0.869 5653
y_4	0.833 3333
y_5	0.800 0000
y_6	0.769 2307
y_7	0.740 7407
y_8	0.714 2857
y_9	0.689 6552
y_{10}	0.666 6667
y_{11}	0.645 1613
y_{12}	0.625 0000
y_{13}	0.606 0606
y_{14}	0.588 2353
y_{15}	0.571 4287
y_{16}	0.555 5556
y_{17}	0.540 5405
y_{18}	0.526 3146
y_{19}	0.512 8205
Σ_1	13.116 6666

$y_{1/2}$	0.975 6097
$y_{3/2}$	0.930 2326
$y_{5/2}$	0.888 8889
$y_{7/2}$	0.851 0638
$y_{9/2}$	0.816 3266
$y_{11/2}$	0.784 3135
$y_{13/2}$	0.754 7169
$y_{15/2}$	0.727 2727
$y_{17/2}$	0.701 7543
$y_{19/2}$	0.677 9661
$y_{21/2}$	0.655 7377
$y_{23/2}$	0.634 9207
$y_{25/2}$	0.615 3846
$y_{27/2}$	0.597 0149
$y_{29/2}$	0.597 7101
$y_{31/2}$	0.563 3804
$y_{33/2}$	0.547 9451
$y_{35/2}$	0.533 3333
$y_{37/2}$	0.519 4806
$y_{39/2}$	0.506 3291
Σ_2	13.861 3816

TRAPEZOID FORMULA

y_0	1.000 0000
y_{20}	0.500 0000

$2\Sigma_1$	26.232 1332
$y_0 + y_{20}$	1.500 0000
Σ	27.732 1332

$$S = \frac{1}{40} \sum \sim 0.693\,303\,33$$

PONCELET'S FORMULA

$2 \sum'_2$	27.722 7632
$\frac{1}{4}(y_0 + y_{20})$	0.375 0000
$-\frac{1}{4}(y_{1/2} + y_{39/2})$	$-$ 0.370 4847
Σ	27.727 2785

$$S = \frac{1}{40} \sum \sim 0.693\,181\,96$$

SIMPSON'S FORMULA

$2 \sum'_1$	26.232 1332
$4 \sum'_2$	55.445 5264
$y_0 + y_{20}$	0.500 0000
Σ	83.177 6596

$$S = \frac{1}{120} \sum \sim 0.693\,147\,16$$

111. Evaluation of definite integrals with variable upper limits. The evaluation of a definite integral with variable upper limit,

$$F(x) = \int_a^x f(x)\,\mathrm{d}x,$$

is required in numerous problems.

Taking the trapezoid formula (43) as basis, the following method can be indicated for finding the approximate value of this integral — not for all x, of course, but for those by which the interval (a, b) is subdivided, i.e. we find:

$$F(a),\ F(x_1),\ F(x_2),\ \ldots,\ F(x_k),\ \ldots,\ F(x_{n-1}),\ F(b).$$

We have by (43):

$$F(x_k) = \int_a^{a+kh} f(x)\,\mathrm{d}x \sim h\left[\frac{y_0 + y_1}{2} + \ldots + \frac{y_{k-1} + y_k}{2}\right], \tag{48}$$

$$F(x_{k+1}) = \int_a^{a+(k+1)h} f(x)\,\mathrm{d}x \sim h\left[\frac{y_0 + y_1}{2} + \ldots + \frac{y_{k-1} + y_k}{2} + \right.$$
$$\left. + \frac{y_k + y_{k+1}}{2}\right] \sim F(x_k) + \frac{1}{2} h(y_k + y_{k+1}). \tag{49}$$

I	II	III	IV	V	VI
k	x_k	y_k	$s_k = y_k + y_{k+1}$	$\sum_{n=1}^{k} s_n$	$F(x_k) = \frac{1}{2} h \sum_{n=1}^{k} s_n$
0	a	y_0		0	0
			$s_1 = y_0 + y_1$		
1	$a + h$	y_1		s_1	$\frac{1}{2} h s_1$
			$s_2 = y_1 + y_2$		
2	$a + 2h$	y_2		$s_1 + s_2$	$\frac{1}{2} h (s_1 + s_2)$
			$s_3 = y_2 + y_3$		
3	$a + 3h$	y_3		$s_1 + s_2 + s_3$	$\frac{1}{2} h (s_1 + s_2 + s_3)$
			$s_4 = y_3 + y_4$		
4	$a + 4h$	y_4		$s_1 + s_2 + s_3 + s_4$	$\frac{1}{2} h (s_1 + s_2 + s_3 + s_4)$
			$s_5 = y_4 + y_5$		
5	$a + 5h$	y_5		$s_1 + s_2 + s_3 + s_4 + s_5$	$\frac{1}{2} h (s_1 + s_2 + s_3 + s_4 + s_5)$
			$s_6 = y_5 + y_6$		
6	$a + 6h$	y_6		$s_1 + s_2 + s_3 + s_4 + s_5 + s_6$	$\frac{1}{2} h (s_1 + s_2 + s_3 + s_4 + s_5 + s_6)$

Having found $F(x_k)$, this formula enables us to pass to the next value $F(x_{k+1}) = F(x_k + h)$.

This computation can be set out as shown on page 287.

112. Graphical methods. These computations can be made graphically, if the graph of the curve $y = f(x)$ is given; in this case, we construct the graph of the integral curve:

$$y = \int_a^x f(x)\, dx = F(x)$$

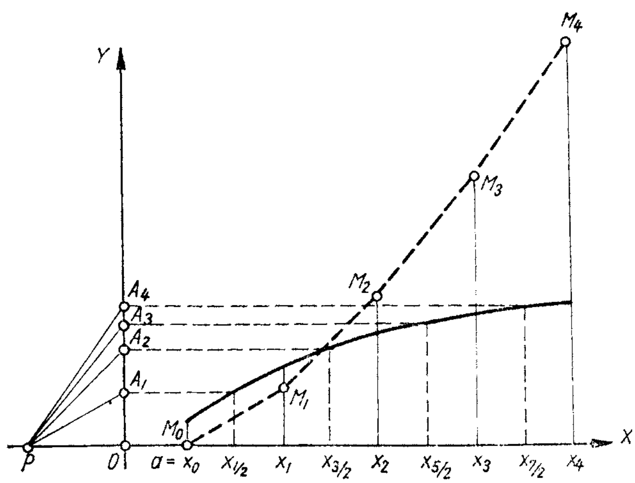

FIG. 149

from the graph of the curve

$$y = f(x). \tag{50}$$

First of all, if we have enough divisions, we can take approximately

$$\frac{s_k}{2} = \frac{y_{k-1} + y_k}{2} = y_{k-1/2}, \tag{51}$$

i.e. if the graph of (50) is drawn, the values $s_k/2$ can be found directly from the figure, as the ordinates of the curve for $x_{k-1/2} = a + 2k - h/2$ (Fig. 149).

We mark on OY the points:

$$A_1\,(y_{1/2}),\quad A_2\,(y_{3/2}),\ldots,\, A_k\,(x_{k-1/2}).$$

We take P to the left of O on OX, so that OP is unity. We draw:

$$PA_1,\, PA_2,\, PA_3,\, \ldots,\, PA_k\,,$$

then draw parallels through M_0, M_1, M_2, $\ldots$, so that

$$M_0M_1 \parallel PA_1,\; M_1M_2 \parallel PA_2,\; M_2M_3 \parallel PA_3,\; \ldots$$

Points $M_0, M_1, M_2, \ldots$ are points of the required approximate integral curve, since it can easily be observed from the figure that

$$\overline{x_1 M_1} = hy_{1/2}, \quad \overline{x_2 M_2} = h(y_{1/2} + y_{3/2}), \quad \overline{x_3 M_3} = h(y_{1/2} + y_{3/2} + y_{5/2})$$

and this shows, by the approximate equation (51), that:

$$\overline{x_k M_k} = h(y_{1/2} + y_{3/2} + \ldots + y_{k-1/2}) =$$
$$= h\left(\frac{y_0 + y_1}{2} + \ldots + \frac{y_{k-1} + y_k}{2}\right) = F(x_k)$$

by (48).

The above construction is for the case when the scale of $F(x)$ is the same as the scale of $f(x)$. If the scale of the area is different, the construction is the same except for making OP of length l instead of unity, where l is the ratio of the scale of $F(x)$ to that of $f(x)$.

An approximate graphical construction of the iterated integral

$$\Phi(x) = \int_a^x \mathrm{d}x \left(\int_a^x f(x)\, \mathrm{d}x \right)$$

is based on rectangle formula (40) [108].

As before, let

$$F(x) = \int_a^x f(x)\, \mathrm{d}x.$$

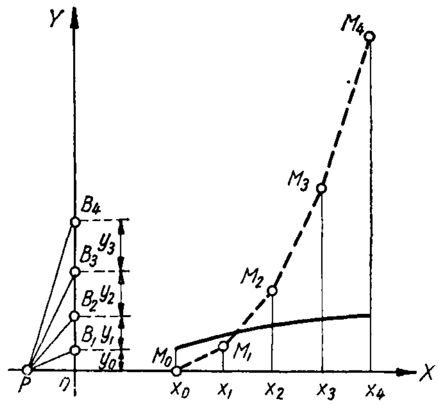

FIG. 150

Taking only the values $x_0, x_1, x_2, \ldots, x_n, \ldots$ of the independent variable x, (40) gives us the approximate equations:

$$F(x_1) \sim hy_0,\ F(x_2) \sim h(y_0 + y_1),\ \ldots,\ F(x_k) \sim h\,(y_0 + y_1 + \ldots + y_{k-1}).$$

Applying the same formula to $\Phi(x)$, we have:

$$\Phi\,(x_k) = h\,[F(x_0) + F(x_1) + \ldots + F(x_{k-1})] \sim$$
$$\sim h^2\,[y_0 + (y_0 + y_1) + \ldots + (y_0 + y_1 + \ldots + y_{k-1})]\,. \qquad (52)$$

This leads us to the following construction for the ordinate $\Phi(x_k)$ (Fig. 150): taking P as before, we cut off on OY:

$$\overline{OB_1} = y_0, \quad \overline{B_1B_2} = y_1, \quad \overline{B_2B_3} = y_2, \ldots, \ \overline{B_{k-1}B_k} = y_{k-1}, \ldots$$

We join:

$$PB_1, PB_2, \ldots, PB_k, \ldots,$$

and take points:

$$M_0, M_1, M_2, \ldots, M_k, \ldots,$$

so that

$$M_0M_1 \parallel PB_1,\ M_1M_2 \parallel PB_2,\ M_2M_3 \parallel PB_3,\ \ldots$$

These points are points of the required approximate curve, drawn, however, on the invariable scale $(1 : h)$, since it is clear from the construction that

$$\overline{x_1 M_1} = hy_0, \quad \overline{x_2 M_2} = hy_0 + h(y_0 + y_1), \ldots, \quad \overline{x_k M_k} =$$
$$= hy_0 + h(y_0 + y_1) = \ldots + h(y_0 + y_1 + \ldots + y_{k-1}) \sim \frac{\Phi(x_k)}{h}$$

by (52). If $\overline{PO}$ is l instead of unity, the curve drawn gives the ordinate of the curve $\Phi(x)$, changed in the ratio $1 : lh$.

It should be mentioned that the accuracy of the above constructions is not great, for all their convenience, and they are only useful for fairly rough calculations.

113. Areas under rapidly oscillating curves. We mentioned in [110] above that the successful application of the various approximate formulae for evaluating definite integrals depends on dividing the curve, the area under which is to be found, into sections in which it has a smooth shape.

This is a very awkward requirement in the case of curves that behave irregularly, with frequent oscillations. If the area under such a curve is to be found in accordance with the above rule, too many subdivisions have to be introduced, and the computation becomes very involved.

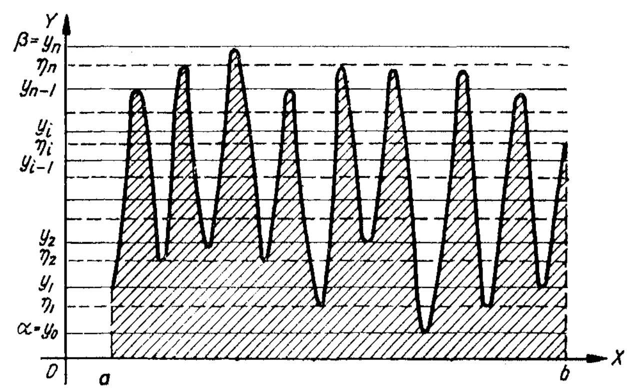

FIG. 151

A second method is useful in these cases, and consists in dividing the area into strips parallel to OX, instead of OY. To find the approximate area under the curve of Fig. 151, we mark off on OY the least and greatest ordinates α and β of the curve, and divide the interval (α, β) into n parts with the points:

$$y_0 = \alpha, y_1, \ldots, y_{i-1}, y_i, \ldots, y_{n-1}, y_n = \beta.$$

We draw the parallels to OX through these points, thus dividing the total area into strips, made up of separate pieces; an approximate expression for the area of the ith strip can be found by taking the product of its base $(y_i - - y_{i-1})$ and the sum of the lengths l_i of the segments cut off on any line

$$y = \eta_i \quad (y_{i-1} < \eta_i < y_i),$$

contained within the area in question; this sum can be found directly from the figure. If l_i denotes the sum, we get an approximate expression for the required area S of the form:

$$y_0(b - a) + (y_1 - y_0)\, l_1 + (y_2 - y_1)\, l_2 + \ldots + (y_n - y_{n-1})\, l_n ,$$

which increases in accuracy, the greater the number of divisions and the sharper the oscillations of the curve.

Suitable development of the basic idea of this method leads to the concept of Lebesgue integrals, which is far more general than the concept described above of Riemann integrals [94, 116].

§ 11. Further remarks on definite integrals

114. Preliminary concepts. The last paragraphs of the present chapter are devoted to the rigorous analytic consideration of the concept of integral, and in fact we shall prove below that a sum of the form

$$\sum_{k=1}^{n} f(\xi_k)\,(x_k - x_{k-1})$$

has a definite limit not only in the case of a continuous function. To do this, we have to introduce some new concepts regarding discontinuous functions. Let $f(x)$ be defined in a certain finite interval (a, b). We shall consider only bounded functions, i.e. those with absolute values less than a given positive number throughout the interval. To be precise, *a function $f(x)$ is said to be bounded in the interval (a, b), if there exists a positive number M, such that we have for all x of the interval :*

$$|f(x)| \leqslant M .$$

If $f(x)$ is continuous, it attains a greatest and least value in the interval, as already remarked [35], and thus it is obviously also bounded. Discontinuous functions, on the other hand, can be unbounded as well as bounded. We shall only consider bounded discontinuous functions in future. We suppose, for instance, that $f(x)$ has a graph as shown in

Fig. 152. We have a discontinuity at $x = c$, and the value of the function at this point, i.e. $f(c)$, has to be defined by means of some supplementary condition. The function is continuous at the remaining points of the interval, including the ends a and b. Also, as x tends to c from smaller values, i.e. from the left, the ordinate $f(x)$ tends to a definite limit, denoted geometrically by $\overline{NM_1}$. Similarly, if x tends to c from larger values, i.e. from the right, $f(x)$ again tends to a definite limit, denoted by $\overline{NM_2}$, this latter limit being different from the previous, left-hand limit. The left-hand limit is usually denoted by $f(c - 0)$, and the right-hand limit by $f(c + 0)$ [32]. The above very

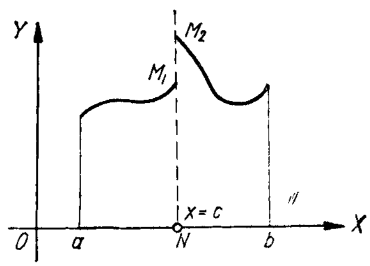

FIG. 152

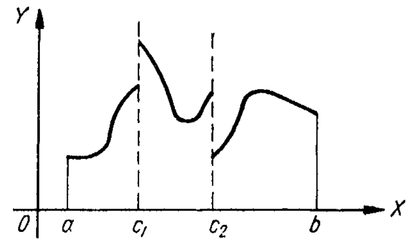

FIG. 153

simple type of *discontinuity, where a finite determinate limit exists both on the left and on the right, is usually called a discontinuity of the first kind.* The value of the function at the point itself, $x = c$, i.e. $f(c)$, will in general differ both from $f(c - 0)$ and $f(c + 0)$, and requires supplementary definition. If a function is continuous in an interval (a, b) including the ends, with the exception of a finite number of points where it has discontinuities of the first kind, the graph consists of a finite number of curves that are continuous as far as their ends, together with distinct points at the places where the discontinuities occur (Fig. 153). Such a function is evidently bounded throughout the interval, in spite of its discontinuities. At the same time, of course, functions with more complicated discontinuities can also be bounded.

We shall often be considering, further on, the set of all the values taken by a function $f(x)$ in a given interval of variation of the independent variable. If the function is bounded in the interval, the set of its values in the interval is bounded above and below, and hence the set has a strict upper and a strict lower bound [39]. If, for instance, $f(x)$ is continuous in the (closed) interval, it attains a greatest and a least value in the interval, as we know from [35]. In this case, the

greatest and least values of the function are also the strict upper and lower bounds of the values of $f(x)$ in the interval. We take another example. If $f(x)$ is an increasing function, it takes its greatest value at the right-hand end of the interval, and its least value at the left. As in the previous case, these values are the strict upper and lower bounds of the values of $f(x)$. In both examples, the strict bounds of the values of the function are particular values of the function, i.e. they themselves belong to the set of values of the function. In the more complex case of discontinuous functions, the strict bounds of the values of the function may not themselves be values of the function, i. e. they may not belong to the set of values of the function.

Let M and m be the strict upper and lower bounds of the values of $f(x)$ in the interval (c, d), i.e. for $c \leqslant x \leqslant d$, where obviously, $m \leqslant M$. We take a new interval (c', d'), which is only part of (c, d). Let M' and m' be the *strict* upper and lower bounds of the values of $f(x)$ in the new interval (c', d'). Since the set of all the values of $f(x)$ in (c', d') must always be included among the values of $f(x)$ in the wider interval (c, d), we can say that $M' \leqslant M$ and $m' \geqslant m$, i.e. *if the interval of variation of x is replaced by a part of the interval, the strict upper bound of the values of function $f(x)$ cannot increase, and the strict lower bound cannot decrease.* This fact has great importance for us later on.

115. Darboux's theorem. Let $f(x)$ be bounded in the interval (a, b), and let m and M be the strict lower and upper bounds of its values in the interval. We divide the interval with points corresponding to the values of x:

$$a = x_0 < x_1 < x_2 < \ldots < x_{k-1} < x_k \ldots < x_{n-1} < x_n = b,$$

and introduce the lengths of the sub-intervals $\delta_k = x_k - x_{k-1}$ $(k = 1, 2, \ldots, n)$. Let $x = \xi_k$ belong to the sub-interval (x_{k-1}, x_k) $(k = 1, 2, \ldots, n)$. We form the sum of the products:

$$\sum_{k=1}^{n} f(\xi_k)\,\delta_k. \tag{1}$$

The value of this sum depends, firstly on the method of division of the interval (a, b), and secondly, on the choice of $x = \xi_k$ in each sub-interval. Our problem is to investigate the limit of the sum when the number n of sub-intervals increases indefinitely and the greatest of the lengths δ_k tends to zero. We have to decide in what cases it is

possible to speak of this limit, i.e. decide for what functions $f(x)$ the sum (1) tends to a definite limit, independently of the method of division of the interval and of the choice of ξ_k.

We consider $f(x)$ in each sub-interval, and let M_k and m_k be the strict upper and lower bounds of its values in (x_{k-1}, x_k). We replace $f(\xi_k)$ in the terms of sum (1) by M_k or m_k.

We thus arrive at the following two sums:

$$S = \sum_{k=1}^{n} M_k \delta_k, \tag{2}$$

$$s = \sum_{k=1}^{n} m_k \delta_k, \tag{3}$$

with the inequality following directly from the definition of strict bounds:

$$m_k \leqslant f(\xi_k) \leqslant M_k,$$

whence we have, since δ_k is positive:

$$s \leqslant \sum_{k=1}^{n} f(\xi_k)\, \delta_k \leqslant S. \tag{4}$$

We consider sums S and s in more detail, then return to the more general sum (1). In accordance with the remark of the previous article, M_k and m_k always satisfy the inequality

$$m \leqslant m_k \leqslant M_k \leqslant M$$

and in addition, obviously:

$$\sum_{k=1}^{n} \delta_k = \sum_{k=1}^{n} (x_k - x_{k-1}) = b - a.$$

It follows directly that

$$m\delta_k \leqslant M_k \delta_k \leqslant M\delta_k \quad \text{and} \quad m\delta_k \leqslant m_k \delta_k \leqslant M\delta_k,$$

whence we get by summing over k:

$$m(b - a) \leqslant \sum_{k=1}^{n} M_k \delta_k \leqslant M(b - a) \text{ and}$$

$$m(b - a) \leqslant \sum_{k=1}^{n} m_k \delta_k \leqslant M(b - a),$$

i.e. sums S and s always lie between the bounds $m(b - a)$ and $M(b - a)$ for any division of the interval.

If we take all the possible divisions of the interval (a, b), we get infinite sets of values for both sums (2) and (3). It follows from what has just been said that both these sets are bounded, and hence, both have strict upper and lower bounds.

We examine the sum S more closely, and assume for the present that all the values of $f(x)$ are positive. In this case, all the terms of S are also positive. We suppose that the interval (a, b) is divided into subintervals δ_k in a definite way, so that S has a definite value.

We make a further division of the sub-interval δ_k.† For instance, let a given δ_k be divided into three parts: $\delta_k^{(1)}$, $\delta_k^{(2)}$, $\delta_k^{(3)}$, and let the corresponding strict upper bounds of $f(x)$ be $M_k^{(1)}$, $M_k^{(2)}$, and $M_k^{(3)}$.

By the remark of the previous article, these strict upper bounds are never greater than the strict upper bound for the full sub-interval δ_k, i.e.

$$M_k^{(1)},\ M_k^{(2)},\ \text{and}\ M_k^{(3)} \leqslant M_k \tag{5}$$

and in addition, obviously:

$$\delta_k^{(1)} + \delta_k^{(2)} + \delta_k^{(3)} = \delta_k. \tag{6}$$

After the further division of δ_k into three parts, the term $M_k\,\delta_k$ of S is replaced by three terms:

$$M_k^{(1)}\,\delta_k^{(1)},\ M_k^{(2)}\,\delta_k^{(2)}\ \text{and}\ M_k^{(3)}\,\delta_k^{(3)},$$

and we have by (5) and (6):

$$M_k^{(1)}\,\delta_k^{(1)} + M_k^{(2)}\,\delta_k^{(2)} + M_k^{(3)}\,\delta_k^{(3)} \leqslant M_k\,\delta_k \tag{7}$$

i.e. *if we start from a definite division of the interval, then further sub-divide individual sub-intervals δ_k, the sum S can only diminish,* or more precisely, *cannot increase.* Instead of considering as a whole the new sum that we get by sub-dividing individual sub-intervals δ_k, we shall later be considering only part of its terms, i.e. we shall be discarding certain terms of the new sum. Since all the terms are positive, the discarding of certain terms can only decrease the value of the complete sum, i.e. the new sum will become *a fortiori* not greater than the original sum S, which we had before sub-division of the sub-intervals δ_k.

We have compared the two values of S corresponding to two types of division of the interval (a, b), the second type of division being obtained by further subdivision of the first, in such a way that all the points of division of the first were preserved in the second.

† For brevity, we use δ_k to denote both the sub-interval and its length.

When the two values of S are compared for two arbitrary methods of division of (a, b), there will in general be no simple relationship between them. But it is clear that, if the lengths of sub-intervals δ_k are sufficiently small for both methods of division, the values of S will be close together in magnitude. More precisely, it can be shown that, on indefinite increase in the number n of divisions, and on indefinite decrease of the greatest of the increments δ_k, S tends to a definite limit, independently of the method of division of the interval (a, b).

We now proceed to the proof of this useful proposition.

We take all the possible values of S, obtained by all the possible divisions of (a, b). Let L be the strict lower bound of this set of values of S which is bounded above and below. We show that L is the abovementioned limit of S.

By the definition of strict lower bound, we have $L \leqslant S$ for all S. To prove our assertion that L is the limit of S, we have to show that, for any given positive ε, there exists an η such that all S are less than $(L + \varepsilon)$ for all δ_k less than η.

It follows from the definition of L as the strict lower bound of values of S, that there exists a fully defined method (I) of division of (a, b) into sub-intervals δ_k, such that the corresponding value of S, say S', is less than $(L + \varepsilon/2)$. Let p be the number of points of division of the total interval (a, b) according to method (I). We take any other method (II) of division of (a, b), and we let δ_k be the lengths of the sub-intervals (x_{k-1}, x_k) as usual. We separate all the increments δ_k into two classes.

We put those that are wholly contained in one of the sub-intervals δ_k' in the first class, δ_k' being obtained by method (I), and we put those that run into more than one δ_k' in the second class. Let σ_l be the lengths of the sub-intervals of the first class, and τ_m the lengths of those of the second class. Further, let μ_l and ν_m be the strict upper bounds of $f(x)$ in the sub-intervals σ_l and τ_m. Subscripts l and m run through certain integers that are of no interest to us, and in future, when summing over these subscripts, we shall not indicate the limits of summation, it being assumed that summation takes place over all the sub-intervals of the first or second class. When separating the δ_k into two classes, we can at the same time separate the total sum S obtained by method (II) of division into two sums:

$$S = S_1 + S_2 ,$$

where

$$S_1 = \sum \mu_l \sigma_l; \quad S_2 = \sum \nu_m \tau_m .$$

Each σ_l forms part of a δ'_k of the first method (I) of division, whereas all the σ_l do not fill up all the δ'_k, i.e. S_1 can be found from S' by sub-dividing the δ'_k and discarding certain terms. We can therefore say, by what was proved above, that S_1 is not greater than S', i.e. we can write, since $S' < L + \varepsilon/2$ ε,

$$S_1 < L + \frac{\varepsilon}{2}. \tag{7_1}$$

We now consider the second sum S_2. The sub-intervals τ_m run into more than one (at least two) δ_k of method of division (I), i.e. τ_m covers at least one point of division of (a, b) in accordance with method (I), so that the number of terms in S_2 is not greater than the number p of points of division by method (I), where p is a definite positive integer. The factors ν_m do not exceed the strict upper bound M of $f(x)$ in the whole of (a, b). If τ denotes the greatest of the τ_m, every term of S_2 is not greater than $M\tau$, and hence, we have for the whole sum:

$$S_2 \leqslant M\tau p. \tag{8}$$

We now take η equal to $\varepsilon/2Mp$ and show that it satisfies the condition stated above. We therefore suppose that all the δ_k of method (II) of division satisfy the inequality:

$$\delta_k \leqslant \frac{\varepsilon}{2Mp}. \tag{9}$$

Since the τ_m are certain of the δ_k, we have $\tau_m \leqslant \varepsilon/2Mp$, i.e.

$$\tau \leqslant \frac{\varepsilon}{2Mp},$$

and (8) gives us for S_2:

$$S_2 \leqslant \frac{\varepsilon}{2}. \tag{10}$$

On adding inequalities (7_1) and (10), we get for the total sum S:

$$S < L + \frac{1}{2}\varepsilon + \frac{1}{2}\varepsilon = L + \varepsilon.$$

We thus have the inequality for the sum S:

$$L \leqslant S < L + \varepsilon,$$

for any method of division of the interval (a, b), provided only that the lengths of the sub-intervals satisfy inequality (9).

In view of the arbitrary smallness of the given positive ε, we conclude from this that L is in fact the limit of sum S.

We have assumed in the above discussion that all the values of $f(x)$ are positive. If this is not the case, we can always add a positive con-

stant A to the bounded function $f(x)$, such that the new function $\psi(x) = f(x) + A$ is positive. Our proposition can now be considered proved for the new function $\psi(x)$, i.e. S has a definite limit for the new function. Since the strict upper bound of $\psi(x)$ is evidently $M_k + A$ in (x_{k-1}, x_k), using the previous notation, we see that the sum for $\psi(x)$ has the form:

$$\sum_{k=1}^{n} (M_k + A)\, \delta_k = \sum_{k=1}^{n} M_k \delta_k + A \sum_{k=1}^{n} \delta_k = \sum_{k=1}^{n} M_k \delta_k + A(b-a),$$

where, as above, M_k is the strict upper bound of $f(x)$ in δ_k.

We now consider the formula obtained:

$$\sum_{k=1}^{n} (M_k + A)\, \delta_k = \sum_{k=1}^{n} M_k \delta_k + A(b-a).$$

The left-hand sum has a definite limit, as remarked above, equal to the strict lower bound of the values of the left-hand sum.

We have two terms on the right-hand side, one of which, $A(b-a)$, is a definite number, so that we can say that the other term $\sum_{k=1}^{n} M_k\, \delta_k$ also has a definite limit, equal to the strict lower bound of the set of values of this sum.

We have thus shown that the sum S has a determinate limit L for any bounded function $f(x)$ in a finite interval. It can similarly be shown that the sum (3) also tends to a determinate limit l on indefinite decrease of the greatest of the δ_k. This number l is the strict upper bound of all the possible values of the sum s for all the possible divisions of interval (a, b). Moreover, noting that $m_k \leqslant M_k$, and comparing expressions (2) and (3) for sums S and s, we see that, given the same method of division, we always have $s \leqslant S$. We therefore get the same inequality for the limits, i.e. $l \leqslant L$. We formulate the result obtained as the following theorem, due to the French mathematician Darboux:

DARBOUX'S THEOREM. *The sums s and S tend to definite limits l and L, where $l \leqslant L$, for any function bounded in the interval (a, b), on indefinite increase in the number of sub-divisions and indefinite decrease of the greatest of the δ_k.*

We said above that l is the upper bound of values of s and L the lower bound of values of S. Having proved that $l \leqslant L$, we can therefore say that $s \leqslant S$, whatever the method of division used for forming s and S.

116. Functions integrable in Riemann's sense. If we now turn to the general sum:

$$\sum_{k=1}^{n} f(\xi_k)\,\delta_k \qquad (\delta_k = x_k - x_{k-1}) \tag{11}$$

it seems that we cannot yet assert the existence of a limit in the case of any bounded function $f(x)$.

There is some uncertainty about the magnitude of the factor $f(\xi_k)$, since ξ_k can be chosen in any manner from the sub-interval (x_{k-1}, x_k). It is due to this fact that the sum (1) does not always have a definite limit. We can suppose, for instance, that the limits l and L, referred to in Darboux's theorem, are not identical, i.e. that $l < L$. By the definition of strict upper and lower bounds, we can choose ξ_k, on the one hand, so that $f(\xi_k)$ is as close as desired to m_k, and on the other, so that $f(\xi_k)$ is as close as desired to M_k. The sum (1) will be as close as desired to the value of the corresponding sum s in the first case, and to that of S in the second case. We can thus, by suitable choice of ξ_k, make (1) as close as desired either to l (the limit of s) or to L (the limit of S), on indefinite decrease of the δ_k. Since l differs from L by hypothesis, it is evident that (11) has no definite limit, on indefinite increase of n and indefinite decrease of the greatest of the δ_k. We have thus shown that (11) *has no definite limit, if $l < L$.*

We now show that, if $l = L$, (1) has a definite limit, equal to $l = L$. We have, in fact, by the definition of strict upper and lower bound, $m_k \leqslant f(\xi_k) \leqslant M_k$, and we can therefore write:

$$\sum_{k=1}^{n} m_k\,\delta_k \leqslant \sum_{k=1}^{n} f(\xi_k)\,\delta_k \leqslant \sum_{k=1}^{n} M_k\,\delta_k.$$

The extreme terms of this inequality have the common limit $l = L$, on indefinite decrease of the greatest of the δ_k, and hence the sum $\sum\limits_{k=1}^{n} f(\xi_k)\,\delta_k$ must tend to the same limit, for any choice of ξ_k. As we know, the limit of this sum is referred to as the definite integral of $f(x)$ over (a, b), and if this limit exists, the function is said to be *integrable* in Riemann's sense, or simply, integrable. Different definitions to the above are given to the definite integral in certain cases, and of course, the condition for integrability is then different. We speak of integrability in Riemann's sense (Riemann was a mid-nineteenth century German mathematician) so as to distinguish his method of forming the definite integral from other methods. Since we shall only

be concerned in future with Riemann integrals, we shall not enlarge on this remark, and functions integrable in Riemann's sense will simply be referred to as integrable.

It follows from the above, that *a necessary and sufficient condition for the integrability of $f(x)$ consists in the coincidence of the limits l and L of sums s and S, i.e. in the fact that the difference between these sums:*

$$\sum_{k=1}^{n} (M_k - m_k)\, \delta_k\,, \tag{12}$$

tends to zero on indefinite increase of n and on indefinite decrease of the greatest of the increments δ_k. We investigate some classes of function, for which the above condition is satisfied, i.e. some classes of integrable function.

The sum (12) consists of non-negative terms, and its magnitude is not less than $L - l$, since L is the strict lower bound of sum (2), and l the strict upper bound of (3). We can say from this, and from Darboux's theorem, that a necessary and sufficient condition for integrability, i.e. for coincidence of l and L, may be expressed as follows: *for any given positive ε, there exists a division of the interval (a, b), such that the sum* (12) *is less than ε.*

1. If $f(x)$ is continuous in (a, b) (including the ends), it is uniformly continuous in the interval. Furthermore, it attains a least value m_i and a greatest value M_i in each sub-interval δ_i. By the uniform continuity of $f(x)$, the positive differences $M_i - m_i$ will be less than any positive ε on indefinite decrease of the greatest of the increments δ_i, and we shall have for the total positive sum (12):

$$0 \leqslant \sum_{i=1}^{n} (M_i - m_i)\, \delta_i \leqslant \sum_{i=1}^{n} \varepsilon\delta_i = \varepsilon(b - a)\,.$$

Hence it follows, by the arbitrary smallness of ε, that (12) tends to zero, i.e. *every continuous function is integrable.*

2. We now take $f(x)$ bounded, with a finite number of discontinuities. We assume, for clarity, that it has one discontinuity at $x = c$ inside (a, b).

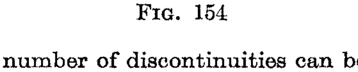

FIG. 154

The case of any finite number of discontinuities can be considered in exactly the same way. Since the limit $L - l$ of (12) does not depend on the method of division of (a, b), we can use whatever method of division we like for our discussion, with the only proviso that all the increments δ_i tend to zero. We isolate the point c from (a, b) by means of a small sub-interval (a_1, b_1) (Fig. 154) such that c is an interior point. The sub-interval is defined more accurately later.

We have $|f(x)| < N$, since $f(x)$ is bounded, i.e. we have all $M_i < N$ and all $m_i > -N$, so that:

$$0 \leqslant M_i - m_i \leqslant 2N. \tag{13}$$

Let ε be any small given positive number. We choose (a_1, b_1) such that

$$2N(b_1 - a_1) < \varepsilon. \tag{14}$$

We shall assume that points a_1, b_1 form points of division of (a, b). The sum (12) is now divided into three parts: S_1, corresponding to the interval (a, a_1); S_2, corresponding to (b_1, b); and S_3, corresponding to (a_1, b_1).

Since $f(x)$ is uniformly continuous in (a, a_1), S_1 tends to zero, as in 1. Similarly for S_2, i.e. S_1 and S_2 are less than ε for all sufficiently small δ_i.

To obtain S_3, we have to perform the summation in (12) over those of the δ_k that belong to (a_1, b_1), the sum of these δ_k being evidently equal to $(b_1 - a_1)$. Using (13) also, we have:

$$0 \leqslant S_3 \leqslant \sum' 2N\delta_k = 2N \sum' \delta_k = 2N(b_1 - a_1),$$

where the summation is over the above-mentioned δ_k. By (14), we have $S_3 < \varepsilon$, and the total positive sum (12) is less than 3ε. Hence we can conclude, since ε is arbitrarily small, that the sum tends to zero, i.e. *every bounded function with a finite number of discontinuities is integrable.* We had such a function in the first example of [97].

3. We take the case of $f(x)$ monotonic and bounded in (a, b). We assume for clarity that the function is non-decreasing, i.e. if $c_1 < c_2$, $f(c_1) \leqslant f(c_2)$. We now have in each δ_i, $M_i = f(x_i)$ and $m_i = f(x_{i-1})$. The sum (12) will be:

$$\sum_{i=1}^{n} [f(x_i) - f(x_{i-1})]\,(x_i - x_{i-1}).$$

Let Δ denote the greatest of the $(x_i - x_{i-1})$. By hypothesis, $\Delta \to 0$. Since $f(x_i) - f(x_{i-1}) \geqslant 0$, we can write:

$$0 \leqslant \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})]\,(x_i - x_{i-1}) \leqslant \Delta \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})],$$

i.e.

$$0 \leqslant \sum_{i=1}^{n} [f(x_i) - f(x_{i-1})]\,(x_i - x_{i-1}) \leqslant \Delta\,[f(b) - f(a)],$$

since obviously:

$$\sum_{i=1}^{n} [f(x_i) - f(x_{i-1})] = [f(x_1) - f(a)] + [f(x_2) - f(x_1)] + \cdots$$
$$\cdots + [f(b) - f(x_{n-1})] = f(b) - f(a).$$

We thus see at once that (12) tends to zero, i.e. *every monotonic bounded function is integrable.* We remark that a monotonic function can have an infinite set of discontinuities, so that case (III) is not dealt with by case (II).

We can take as an example the function equal to zero for $0 \leqslant x < 1/2$, equal to $1/2$ for $1/2 \leqslant x < 2/3$, equal to $2/3$ for $2/3 \leqslant x < 3/4$, and so on, and finally equal to 1 for $x = 1$.

The discontinuities of this non-decreasing function are at:

$$x = \frac{1}{2}, \quad \frac{2}{3}, \quad \frac{3}{4}, \quad \frac{4}{5}, \ldots .$$

We remark that every monotonic bounded function must have limits $f(c-0)$ and $f(c+0)$ at every discontinuity $x = c$. This follows directly from the existence of a limit for a monotonic bounded sequence [30].

We have found the condition for integrability on the assumption, throughout, that $f(x)$ is bounded. It can be shown that this is a necessary condition for integrability, i.e. for the existence of a definite limit of the sum (11). If the condition of boundedness is not satisfied, the integral of $f(x)$ over the interval (a, b) can still be defined in certain cases, though no longer as the limit of (11). The integral is then said to be improper. The basic principles of improper integrals were explained in [97]. They are described in more detail in Volume II.

If integration is over an interval that is infinite at one end or the other, the definite integral can again not be conceived directly as the limit of a sum of the form (11), and we again have an improper integral (cf. [98] of Volume II).

117. Properties of integrable functions. The basic properties of integrable functions are easily obtained by using the necessary and sufficient condition for integrability given above.

I. *If $f(x)$ is integrable in (a, b), and we arbitrarily change the value of $f(x)$ at a finite number of points of (a, b), the new function will also be integrable in (a, b), and its integral will not differ from that of $f(x)$.*

We confine ourselves to the case when $f(x)$ is changed at one point, say $x = a$. The new function $\varphi(x)$ coincides with $f(x)$ everywhere except at $x = a$, $\varphi(a)$ being taken arbitrarily. Let m and M be the strict lower and upper bounds of $f(x)$ in (a, b). The strict lower bound of $\varphi(x)$ will evidently be greater than or equal to m, if $\varphi(a) \geqslant m$, and will be $\varphi(a)$ if $\varphi(a) < m$. Similarly, the strict upper bound of $\varphi(x)$ will be less than or equal to M, if $\varphi(a) \leqslant M$, and will be $\varphi(a)$ if $\varphi(a) > M$. Comparing the sum (12) for $f(x)$ and $\varphi(x)$, we note that there can only be a difference in the first term (for $k = 1$). But this first term evidently tends to zero for $f(x)$ and $\varphi(x)$, since $\delta_1 \to 0$ and $(M_1 - m_1)$ is bounded. The sum of the remaining terms, excluding the first, evidently also tends to zero, since $f(x)$ is integrable, and the entire sum (12) must tend to zero in the case of $f(x)$. The integrability of $\varphi(x)$ is proved. The identity of the integrals of $f(x)$ and $\varphi(x)$ is

evident, since we can always take ξ_1 different from a when forming the sum (11), and $f(x)$ and $\varphi(x)$ coincide at all points except $x = a$.

II. *If $f(x)$ is integrable in (a, b), it is integrable in any (c, d) forming part of (a, b).*

We can find the limits l and L of sums s and S on the assumption that c and d are among the points of division of (a, b). The sum (12) for (c, d) is now obtained from the sum (12) for (a, b) simply by subtracting the terms corresponding to (a, c) and (d, b). Since the terms are non-negative, (12) for (c, d) is less than or equal to (12) for (a, b), so that if the latter sum tends to zero [$f(x)$ integrable in (a, b)], the former will certainly also tend to zero, i.e. $f(x)$ is integrable in (c, d). We remark that c can coincide with a, whilst d can coincide with b. The equality is proved as in [94]:

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx \quad (a < c < b).$$

III. *If $f(x)$ is integrable in (a, b), $cf(x)$, where c is any constant, is also integrable in (a, b).*

Taking $c > 0$, for instance, we see that the previous m_k and M_k have to be replaced by cm_k and cM_k for $cf(x)$. The sum (12) only acquires the factor c and will tend to zero as before. Property V of [94] is evidently preserved and is proved as before.

IV. *If $f_1(x)$ and $f_2(x)$ are integrable in (a, b), their sum $\varphi(x) = f_1(x) + f_2(x)$ is also integrable in (a, b).*

Let m_k', M_k', m_k'', M_k'' be the strict lower and upper bounds of $f_1(x)$ and $f_2(x)$ in the sub-interval (x_{k-1}, x_k). Then all the values of $f_1(x)$ in (x_{k-1}, x_k) are greater than or equal to m_k', and all those of $f_2(x)$ are similarly greater than or equal to m_k''. Hence, we have $\varphi(x) \geqslant m_k' + m_k''$ in (x_{k-1}, x_k). We can show similarly that $\varphi(x) \leqslant M_k' + M_k''$ in (x_{k-1}, x_k). We let m_k and M_k denote the strict lower and upper bounds of $\varphi(x)$ in (x_{k-1}, x_k), so that we have $m_k \geqslant m_k' + m_k''$ and $M_k \leqslant M_k' + M_k''$, whence it follows at once that:

$$M_k - m_k \leqslant (M_k' + M_k'') - (m_k' + m_k''),$$

i.e.

$$M_k - m_k \leqslant (M_k' - m_k') + (M_k'' - m_k'').$$

Forming the sum (12) for $\varphi(x)$, we get:

$$0 \leqslant \sum_{k=1}^{n} (M_k - m_k)\,\delta_k \leqslant \sum_{k=1}^{n} (M_k' - m_k')\,\delta_k + \sum_{k=1}^{n} (M_k'' - m_k'')\,\delta_k\,.$$

Both sums on the right tend to zero, since $f_1(x)$ and $f_2(x)$ are integrable by hypothesis. It follows that sum (12) for $\varphi(x)$, i.e.

$$\sum_{k=1}^{n} (M_k - m_k)\, \delta_k,$$

certainly tends to zero, i.e. $\varphi(x)$ is also integrable. The proof is easily extended to the case of the algebraic sum of any finite number of terms. Property VI of [94] is proved as before.

The proofs of the following properties are similar to the above:

V. *The product of two functions integrable in* (a, b) *is also integrable in* (a, b).

VI. *If* $f(x)$ *is integrable in* (a, b), *where its strict lower and upper bounds have the same sign,* $1/f(x)$ *is also integrable in* (a, b).

VII. *If* $f(x)$ *is integrable in* (a, b), *its absolute value* $|f(x)|$ *is also integrable in* (a, b).

The inequality (10) of [95] can be proved as before. The proof of property VII of [95] remains the same, if $f(x)$ and $\varphi(x)$ are integrable functions. The mean value theorem reads: if $f(x)$ and $\varphi(x)$ are integrable in (a, b), and $\varphi(x)$ preserves the same sign in the interval, then

$$\int_a^b f(x)\, \varphi(x)\, \mathrm{d}x = \mu \int_a^b \varphi(x)\, \mathrm{d}x,$$

where μ is a number satisfying $m \leqslant \mu \leqslant M$, and m and M are the strict lower and upper bounds of $f(x)$ in (a, b). In particular:

$$\int_a^b f(x)\, \mathrm{d}x = \mu(b - a).$$

The proof is as before [95]. It can easily be shown, with the aid of this formula, that

$$F(x) = \int_a^x f(t)\, \mathrm{d}t$$

is a continuous function of x, and $F'(x) = f(x)$ for all values of x at which $f(x)$ is continuous. We finally establish the basic formula for the evaluation of integrals of integrable functions. Let $F_1(x)$ be continuous in (a, b), and let its derivative $F_1'(x) = f(x)$ for any x inside (a, b), where $f(x)$ is integrable in (a, b).

We have the basic formula in this case:

$$\int_a^b f(x)\, \mathrm{d}x = F_1(b) - F_1(a).$$

On dividing the interval as usual, and applying the formula for finite increments [63] to each sub-interval (x_{k-1}, x_k), we can write:

$$F_1(x_k) - F_1(x_{k-1}) = F_1'(\xi_k)\,\delta_k = f(\xi_k)\,\delta_k \quad (x_{k-1} < \xi_k < x_k). \quad (15)$$

Further, summing over k and noting that (III of [116]):

$$\sum_{k=1}^{n}[F_1(x_k) - F_1(x_{k-1})] = F_1(b) - F_1(a),$$

we get:

$$F_1(b) - F_1(a) = \sum_{k=1}^{n} f(\xi_k)\,\delta_k.$$

This equation applies for any division of (a, b), in view of the nature of the choice of the ξ_k, which is determined by the formula for finite increments (15). We get the integral instead of the sum on passing to the limit:

$$F_1(b) - F_1(a) = \int_a^b f(x)\,\mathrm{d}x,$$

which it was required to prove. We remark that the values of $f(x)$ at the ends of (a, b) play no part in the definition of the integral, by property I of the present section.

EXERCISES ON CHAPTER III

Find the integrals of the function **1**—**20**:

1. $5a^2x^6$. **2.** $(6x^2 + 8x + 3)$. **3.** $x(x + a)(x + b)$. **4.** $(a + bx^3)^2$.
5. $\sqrt{(2px)}$. **6.** $x^{-1/n}$. **7.** $(nx)^{(1-n)/n}$. **8.** $(a^{2/3} - x^{2/3})^3$.
9. $(\sqrt{x + 1})(x - \sqrt{x} + 1)$.
10. $\dfrac{(x^2 + 1)(x^2 - 2)}{\sqrt[3]{x^2}}$.
11. $(x^m - x^n)^2/\sqrt{x}$. **12.** $(\sqrt{a} - \sqrt{x})^4/\sqrt{ax}$ **13.** $1/(x^2 + 7)$.
14. $1/(x^2 - 10)$. **15.** $1/\sqrt{4 + x^2}$. **16.** $1/\sqrt{8 - x^2}$.
17. $\dfrac{\sqrt{2 + x^2} - \sqrt{2 - x^2}}{\sqrt{4 - x^2}}$
18. $\tan^2 x$. **19.** $\coth^2 x$. **20.** $3^x\,\mathrm{e}^x$.

Evaluate the integrals **21**—**25** by considering them as limits of sums:

21. $\int_a^b \mathrm{d}x$ **22.** $\int_0^T (v_0 + \mathrm{g}t)\,\mathrm{d}t$.

23. $\int_{-2}^{1} x^2\,\mathrm{d}x$. **24.** $\int_0^{10} 2^x\,\mathrm{d}x$ **25.** $\int_1^5 x^3\,\mathrm{d}x$

Find the derivatives with respect to x of the functions **26—29**:

26. $\int_1^x \log t \, dt$.

27. $\int_x^0 \sqrt{1+t^4} \, dt$. **28.** $\int_x^{x^2} e^{-t^2} dt$.

29. $\int_{1/x}^{\sqrt{x}} \cos(t^2) \, dt$.

30. Find the turning points of the function

$$y = \int_0^x \frac{\sin t}{t} \, dt \text{ for } x > 0.$$

By evaluating the corresponding integral evaluate the limits **31—33**:

31. $\lim_{n\to\infty} (1/n^2 + 2/n^2 + \ldots + (n-1)/n^2)$,

32. $\lim_{n\to\infty} [1/(n+1) + 1/(n+2) + \ldots + 1/(n+n)]$,

33. $\lim_{n\to\infty} \frac{1^P + 2^P + \ldots + n^P}{n^{P+1}}$, $(P > 0)$.

Evaluate the integrals **34—45**:

34. $\int_1^2 (x^2 - 2x + 3) \, dx$. **35.** $\int_0^\infty (\sqrt{2x} + \sqrt[3]{x}) \, dx$.

36. $\int_1^4 (1+\sqrt{y})/y^2 \, dy$. **37.** $\int_2^6 \sqrt{x-2} \, dx$.

38. $\int_0^{-3} \frac{dx}{\sqrt{25+3x}}$. **39.** $\int_{-2}^{-3} \frac{dx}{x^2-1}$.

40. $\int_0^1 \frac{x \, dx}{x^2+3x+2}$. **41.** $\int_{-1}^{-1} \frac{y^5 \, dy}{y+2}$

42. $\int_1^0 \frac{dx}{x^2+4x+5}$. **43.** $\int_3^4 \frac{dx}{x^2-3x+2}$.

44. $\int_0^1 \frac{z^3 \, dz}{z^8+1}$. **45.** $\int_{\pi/6}^{\pi/4} \sec^2 x \, dx$.

Find the indefinite integrals of the functions **46—122**:

46. $a/(a-x)$. **47.** $(2x+3)/(2x+1)$. **48.** $(1-3x)/(3+2x)$

49. $x/(a+bx)$. **50.** $(ax+b)/(\alpha x + \beta)$. **51.** $(x^2+1)/(x-1)$.

52. $(x^2+5x+7)/(x+3)$. **53.** $(x^4+x^2+1)/(x-1)$. **54.** $[a + b/(x-a)]^2$.

55. $x/(x+1)^2$. **56.** $b/\sqrt{1-x}$. **57.** $\sqrt{a-bx}$. **58.** $(x+3)/\sqrt{x^2-4}$.
59. $x/(x^2-5)$. **60.** $x/(2x^2+3)$. **61.** $(ax+b)/(a^2x^2+b^2)$.
62. $x/\sqrt{a-x}$. **63.** $x^2/(1+x^6)$. **64.** $x^2/\sqrt{x^6-1}$.
65. $(4+x^2)^{-1}$ arc tan $\frac{1}{2}x$. **66.** $\sqrt{\frac{\text{arc sin } x}{1-x^2}}$.
67. $a\mathrm{e}^{-mx}$. **68.** 4^{2-3x}. **69.** $(\mathrm{e}^x-\mathrm{e}^{-x})$. **70.** $(\mathrm{e}^{x/a}+\mathrm{e}^{-x/a})^2$.
71. $a^{-x}b^{-x}(a^x-b^x)^2$. **72.** $a^{-1/2x}(a^{2x}-1)$. **73.** $x\mathrm{e}^{-(1+x^2)}$.
74. $x\,7^{x^2}$. **75.** $x^{-2}\mathrm{e}^{1/x}$. **76.** $5^{\sqrt{x}}/\sqrt{x}$ **77.** $\mathrm{e}^x/(\mathrm{e}^x-1)$. **78.** $\mathrm{e}^x\sqrt{a-b\mathrm{e}^x}$.
79. $x(5-x^2)^{1/5}$. **80.** $(x^3-1)/(x^4-4x+1)$. **81.** $x^3/(x^8+5)$.
82. $x\mathrm{e}^{-x^2}$. **83.** $[3-\sqrt{2+3x^2}]/(2+3x^2)$. **84.** $(x^3-1)/(x+1)$.
85. $\mathrm{e}^{-1/2x}$. **86.** $(1-\sin x)/(x+\cos x)$. **87.** $(\tan 3x-\cot 3x)/\sin 3x$.
88. $1/(x\log^2 x)$ **89.** $\sec^2 x(\tan^2 x-2)^{-1/2}$. **90.** $\left(2+\frac{x}{2x^2+1}\right)\frac{1}{2x^2+1}$.
91. $a^{\sin x}\cos x$. **92.** $x^2(x^3+1)^{-1/3}$. **93.** $x(1-x^4)^{-1/2}$.
94. $\tan^2 ax$. **95.** $\sin^2\frac{1}{2}x$. **96.** $\sec^2 x(4-\tan^2 x)^{-1/2}$.
97. $\sec(x/a)$. **98.** $x^{-1}(1+\log x)^{1/3}$. **99.** $(x-1)^{-1/2}\tan\sqrt{x-1}$.
100. $x\,\mathrm{cosec}\,x^2$. **101.** $(1+x^2)^{-1}\{\mathrm{e}^{\text{arc tan } x}+x\log(1+x^2)+1\}$.
102. $(\sin x-\cos x)/(\sin x+\cos x)$. **103.** $\left(1-\sin\frac{x}{\sqrt{2}}\right)^2\mathrm{cosec}\frac{x}{\sqrt{2}}$.
104. $x^2/(x^2-2)$. **105.** $(1+x)^2/x(1+x^2)$. **106.** $\sin 2x\cdot\mathrm{e}^{\sin^2 x}$.
107. $(5-3x)(4-3x^2)^{-1/2}$. **108.** $(\mathrm{e}^x+1)^{-1}$.
109. $[(a+b)+(a-b)x^2]^{-1}$, $0<b<a$. **110.** $\mathrm{e}^x/\sqrt{\mathrm{e}^{2x}-1}$.
111. $\sec ax\cdot\mathrm{cosec}\,ax$. **112.** $\sin(2\pi x/T+\Phi_0)$. **113.** $1/[x(4-\log^2 x)]$.
114. $(4-x^2)^{-1/2}$ arc cos $\frac{1}{2}x$. **115.** $\sec^2 x\cdot\mathrm{e}^{-\tan x}$.
116. $\sin x\cos x(2-\sin^4 x)^{-1/2}$. **117.** $\sec^2 x\cdot\mathrm{cosec}^2 x$.
118. $(1-x^2)^{-1/2}(x+\text{arc sin } x)$. **119.** $\sec x\tan x(\sec^2 x+1)^{-1/2}$.
120. $\cos 2x/(4+\cos^2 2x)$. **121.** $1/(1+\cos^2 x)$.
122. $(1+x^2)^{-1/2}[\log\{x+\sqrt{x^2+1}\}]^{1/2}$.

Find the indefinite integrals of the functions **123—140** by changing the variable of integration:

123. (a) $x^{-1}(x^2-2)^{-1/2}$, ($x=t^{-1}$); (b) $(\mathrm{e}^x+1)^{-1}$, ($x=-\log t$);
(c) $x(5x^2-3)^7$, ($5x^2-3=t$); (d) $x/\sqrt{x+1}$, ($t=\sqrt{x+1}$);
(e) $\cos x/\sqrt{1+\sin^2 x}$, ($t=\sin x$).
124. $x(2x+5)^{10}$. **125.** $(1+x)/(1+\sqrt{x})$. **126.** $1/x\sqrt{2x+1}$.
127. $(\mathrm{e}^x-1)^{-1/2}$. **128.** $x^{-1}(\log 2x/\log 4x)$. **129.** $(1-x^2)^{-1/2}(\text{arc sin } x)^2$.
130. $\mathrm{e}^{2x}(\mathrm{e}^x+1)^{-1/2}$. **131.** $\sin^3 x(\cos x)^{-1/2}$. **132.** $1/x\sqrt{1+x^2}$.
133. $x^2/\sqrt{1-x^2}$. **134.** $x^3/\sqrt{2-x^2}$. **135.** $x^{-1}\sqrt{x^2-a^2}$.
136. $x^{-1}(x^2-1)^{-1/2}$. **137.** $x^{-1}(x^2+1)^{1/2}$. **138.** $x^{-2}(4-x)^{-1/2}$.
139. $\sqrt{1-x^2}$. **140.** $x(1-x)^{-1/2}$, (Put $x=\sin^2 t$).

Find the indefinite integrals of the functions **141—160** by integrating by parts:

141. $\log x$. **142.** $\text{arc}\tan x$. **143.** $\text{arc}\sin x$. **144.** $x \sin x$. **145.** $x \cos 3x$.
146. xe^{-x}. **147.** $x \cdot 2^{-x}$. **148.** $x^2 e^{3x}$. **149.** $(x^2 - 2x + 5)\, e^{-x}$.
150. $x^3 e^{-x/3}$. **151.** $x \sin x \cos x$. **152.** $(x^2 + 5x + 6) \cos 2x$. **153.** $x^2 \log x$.
154. $\log^2 x$. **155.** $x^{-3} \log x$. **156.** $x^{-1/2} \log x$. **157.** $x\,\text{arc} \tan x$. **158.** $x\,\text{arc} \sin x$.
159. $\log [x + \sqrt{1 + x^2}]$. **160.** $x \cdot \operatorname{cosec}^2 x$.

Find the indefinite integrals of the functions **161—218**:

161. $(x^2 + 2x + 5)^{-1}$. **162.** $(x^2 + 2x)^{-1}$. **163.** $(3x^2 - x + 1)^{-1}$.
164. $x\,(x^2 - 7x + 13)^{-1}$. **165.** $(3x - 2)/(x^2 - 4x + 5)$.
166. $(x - 1)^2/(x^2 + 3x + 4)$. **167.** $x^2/(x^2 - 6x + 10)$.
168. $(2 + 3x - 2x^2)^{-1/2}$. **169.** $(x - x^2)^{-1/2}$. **170.** $(x^2 + px + q)^{-1/2}$.
171. $(x^2 + ax + bx + ab)^{-1}$. **172.** $(x^2 - 5x + 9)/(x^2 - 5x + 6)$.
173. $[(x + 1)\,(x + 2)\,(x + 3)]^{-1}$.
174. $(2x^2 + 41\,x - 91)/(x - 1)\,(x + 3)\,(x - 4)$.
175. $(5x^3 + 2)/(x^3 - 5x^2 + 4x)$. **176.** $x^{-1}(x + 1)^{-2}$.
177. $(x^3 - 1)/(4x^3 - x)$. **178.** $(x^4 - 6x^3 + 12x^2 + 6)/(x^3 - 6x^2 + 12x - 8)$
179. $(5x^2 + 6x + 9)/(x - 3)^2\,(x + 1)^2$.
180. $(x^2 - 8x + 7)/(x^2 - 3x - 10)^2$. **181.** $(2x - 3)/(x^2 - 3x + 2)^3$.
182. $(x^3 + x + 1)/x\,(x^2 + 1)$. **183.** $x^4/(x^4 - 1)$.
184. $(x^2 - 4x + 3)^{-1}\,(x^2 + 4x + 5)^{-1}$. **185.** $\cos^3 x$. **186.** $\sin^5 x$.
187. $\sin^2 x \cos^3 x$. **188.** $\sin^3 \tfrac{1}{2}x \cdot \cos^5 \tfrac{1}{2}x$. **189.** $\cos^5 x \cdot \operatorname{cosec}^3 x$.
190. $\sin^4 x$. **191.** $\sin^2 x \cos^2 x$. **192.** $\sin^2 x \cos^4 x$. **193.** $\cos^6 3x$.
194. $\operatorname{cosec}^4 x$. **195.** $\sec^6 x$. **196.** $\cos^2 x \cdot \operatorname{cosec}^6 x$. **197.** $\operatorname{cosec}^2 x \cdot \sec^4 x$.
198. $\operatorname{cosec}^5 x \cdot \sec^3 x$. **199.** $\operatorname{cosec} \tfrac{1}{2}x \cdot \sec^3 \tfrac{1}{2}x$.
200. $\sin (x + \tfrac{1}{4}\pi \operatorname{cosec} x \cdot \sec x$. **201.** $\operatorname{cosec}^5 x$. **202.** $\sec^5 4x$.
203. $\tan^2 5x$. **204.** $\cot^3 x$. **205.** $\cot^4 x$. **206.** $\tan^3 (x/3) + \tan^4 (x/4)$.
207. $x \sin^2 x^2$. **208.** $\sin^5 x \cos^{1/3} x$. **209.** $\operatorname{cosec}^{1/2} x \sec^{3/2} x$. **210.** $\cot^{1/2} x$.
211. $\sin 3x \cos 5x$. **212.** $\sin 10x \sin 15x$. **213.** $\cos \tfrac{1}{2} x \cos \tfrac{1}{3} x$.
214. $\sin \tfrac{1}{3} x \cos \tfrac{2}{3} x$. **215.** $\cos (ax + b) \cos (ax - b)$.
216. $\sin \omega x \sin (\omega x + \varphi)$. **217.** $\cos x \cos^2 3x$. **218.** $\sin x \sin 2x \sin 3x$.

Find the indefinite integrals of the functions **219–227** by making the change of variable $\tan \tfrac{1}{2} x = t$ and integrating with respect to t:

219. $(3 + 5 \cos x)^{-1}$. **220.** $(\sin x + \cos x)^{-1}$. **221.** $\cos x/(1 + \cos x)$.
222. $\sin x/(1 - \sin x)$. **223.** $(8 - 4 \sin x + 7 \cos x)^{-1}$.
224. $(\cos x + 2 \sin x + 3)^{-1}$. **225.** $(3 \sin x + 2 \cos x)/(2 \sin x + 3 \cos x)$
226. $(1 + \tan x)/(1 - \tan x)$. **227.** $(1 + 3 \cos^2 x)^{-1}$.

Find the indefinite integrals of the functions **228—233** by changing the variable of integration by a substitution involving hyperbolic functions:

228. $(3 - 2x - x^2)^{1/2}$. **229.** $\sqrt{2 + x^2}$. **230.** $x^2 (9 + x^2)^{-1/2}$.
231. $\sqrt{x^2 - 2x + 2}$. **232.** $\sqrt{x^2 - 4}$. **233.** $\sqrt{x^2 + x}$.

Examine for convergence (and in the cases when the integrals are convergent evaluate) the definite integrals **234—254**:

234. $\int\limits_0^1 \frac{dx}{\sqrt{x}}$ **235.** $\int\limits_{-1}^2 \frac{dx}{x}$. **236.** $\int\limits_0^1 \frac{dx}{x^p}$ **237.** $\int\limits_0^3 \frac{dx}{(x-1)^2}$

238. $\int\limits_0^1 \frac{dx}{\sqrt{1 - x^2}}$ **239.** $\int\limits_1^\infty \frac{dx}{x}$ **240.** $\int\limits_1^\infty \frac{dx}{x^2}$ **241.** $\int\limits_1^\infty \frac{dx}{x^p}$

242. $\int\limits_{-\infty}^\infty \frac{dx}{1 + x^2}$ **243.** $\int\limits_{-\infty}^\infty \frac{dx}{x^2 + 4x + 9}$ **244.** $\int\limits_0^\infty \sin x \, dx$.

245. $\int\limits_0^{1/2} \frac{dx}{x \log x}$ **246.** $\int\limits_0^{1/2} \frac{dx}{x \log^2 x}$ **247.** $\int\limits_a^\infty \frac{dx}{x \log x} \; (a > 1)$

248. $\int\limits_a^\infty \frac{dx}{x \log^2 x} \; (a > 1)$ **249.** $\int\limits_0^{\frac{1}{2}\pi} \cot x \, dx$.

250. $\int\limits_0^\infty e^{-kx} dx \quad (k > 0)$ **251.** $\int\limits_0^\infty \frac{\arctan x}{x^2 + 1} dx$

252. $\int\limits_2^\infty \frac{dx}{(x^2 - 1)^2}$ **253.** $\int\limits_0^\infty \frac{dx}{x^3 + 1}$ **254.** $\int\limits_0^1 \frac{dx}{x^3 - 5x^2}$

Investigate the convergence of the integrals **255—261**:

255. $\int\limits_0^{100} \frac{dx}{\sqrt[3]{x} + 2\sqrt[4]{x} + x^3}$ **256.** $\int\limits_1^\infty \frac{dx}{2x + \sqrt[3]{x^2 + 1} + 5}$

257. $\int\limits_{-1}^\infty \frac{dx}{x^2 + \sqrt[3]{x^4 + 1}}$ **258.** $\int\limits_0^\infty \frac{x \, dx}{\sqrt{x^5 + 1}}$

259. $\int\limits_0^1 \frac{dx}{\sqrt[3]{1 - x^4}}$ **260.** $\int\limits_1^2 \frac{dx}{\log x}$ **261.** $\int\limits_{\pi/2}^\infty \frac{\sin x}{x^2} dx$

262. Prove that Euler's integral of the first kind *(the Beta function)* $B(p, q)$, defined by the equation

$$B(p, q) = \int\limits_0^1 x^{p-1} (1 - x)^{q-1} dx,$$

is convergent if $p > 0$, $q > 0$.

263. Prove that Euler's integral of the second kind *(the Gamma function)* $\Gamma(p)$, defined by the equation

$$\Gamma(p) = \int_0^\infty x^{p-1} \mathrm{e}^{-x} \mathrm{d}x,$$

is convergent if $p > 0$.

264. Find the area bounded by the parabola $y = 4x - x^2$ and the x-axis.

265. Find the area bounded by the curve $y = \log x$, the x-axis and the line $x = \mathrm{e}$.

266. Find the area bounded by the curve $y = x(x - 1)(x - 2)$ and the x-axis.

267. Calculate the area bounded by the curve $y^3 = x$ and the lines $y = 1$ and $x = 8$.

268. Calculate the area above the x-axis of one arc of the sinusoidal curve $y = \sin x$.

269. Calculate the area bounded by the curve $y = \tan x$, the x-axis and the line $x = \frac{1}{3}\pi$.

270. Find the area bounded by the parabola $xy = m^2$, the lines $x = a$, $x = 3a$ $(a > 0)$ and the x-axis.

271. Find the area bounded by the Witch of Agnesi $y = a^3/(x^2 + a^2)$ and the x-axis.

272. Find the area of the figure bounded by the cubic curve $y = x^3$, the line $y = 8$ and the y-axis.

273. Find the area bounded by the parabolas $y^2 = 2px$ and $x^2 = 2py$.

274. Calculate the area bounded by the parabola $y = 2x - x^2$ and the line $y = -x$.

275. Calculate the area bounded by the line $y = 3 - 2x$ and the parabola $y = x^2$.

276. Find the area of the figure bounded by the parabolas $y = x^2$, $y = \frac{1}{2}x^2$ and the line $y = 2x$.

277. Find the area bounded by the curve $a^2 y^2 = x^2 (a^2 - x^2)$.

278. Find the area enclosed by the curve $(x/5)^2 + (y/4)^{2/3} = 1$.

279. Find the area contained within the astroid $x = a\cos^3 t$, $y = b\sin^3 t$.

280. Find the area bounded by the x-axis and one arc of the cycloid $x = a(t - \sin t)$, $y = a(1 - \cos t)$.

281. Find the area bounded by one branch of the trochoid $x = at - b\sin t$, $y = a - b\cos t$ $(0 < b \leqslant a)$, and the tangent at its lowest point.

282. Find the area bounded by the cardioid $x = a\,(2\cos t - \cos 2t)$, $y = a\,(2\sin t - \sin 2t)$.

283. Find the area of a loop of the folium of Descartes:

$$x = 3at/(1 + t^3); \qquad y = 3at^2/(1 + t^3).$$

284. Find the area of the cardioid $r = a\,(1 + \cos\theta)$.

285. Find the area of one leaf of the curve $r = a\cos 2\theta$.

286. Find the total area enclosed by the curve $r^2 = a^2\sin 4\theta$.

287. Find the area contained within the curve $r = a\sin 3\theta$.

288. Find the area bounded by Pascal's Limaçon $r = 2 + \cos\theta$.

289. Find the area bounded by the parabola $r = a\sec^2\frac{1}{2}\theta$ and the lines $\theta = \frac{1}{4}\pi$ and $\theta = \frac{1}{2}\pi$.

290. Find the area of the ellipse $r = p/(1 + e\cos\theta)$, $(e < 1)$.

291. Find the area bounded by the curve $r = 2a\cos 3\theta$ and lying outside the circle $r = a$.

292. Find the area contained within the curve $x^4 + y^4 = x^2 + y^2$.

293. Calculate the length of the arc of the semi-cubical parabola $y^2 = x^3$ from the origin to the point $(4,8)$.

294. Find the length of the arc of the parabola $y = 2\sqrt{x}$ from $x = 0$ to $x = 1$.

295. Find the length of the part of the curve $y = e^x$ lying between the points $(0,1)$ and $(1, e)$.

296. Find the length of the curve $y = \log x$ from $x = \sqrt{3}$ to $x = \sqrt{8}$.

297. Find the length of the curve $y = \arcsin(e^{-x})$ from $x = 0$ to $x = 1$.

298. Calculate the length of the arc of the curve $x = \log\sec y$ contained between $y = 0$ and $y = \frac{1}{3}\pi$.

299. Find the length of the curve $x = \frac{1}{4}y^2 - \frac{1}{2}\log y$ from $y = 1$ to $y = e$.

300. Find the perimeter of the closed loop of the curve $9ay^2 = x\,(x - 3a)^2$.

301. Find the length of the arc of the evolute of the circle $x = a\,(\cos t + t\sin t)$, $y = a\,(\sin t - t\cos t)$ from $t = 0$ to $t = T$.

302. Find the length of the evolute of the ellipse

$$x = (c^2/a)\cos^3 t, \quad y = (c^2/b)\sin^3 t, \quad (c^2 = a^2 - b^2).$$

303. Find the length of the curve

$$x = a\,(2\cos t - \cos 2t), \quad y = a\,(2\sin t - \sin 2t).$$

304. Find the length of a complete turn of the Archimedean spiral $r = a\,\theta$.

305. Find the length of the perimeter of the cardioid $r = a(1 + \cos\theta)$.

306. Find the length of the part of the parabola $r = a\sec^2\frac{1}{2}\theta$ cut off from the parabola by a vertical line through the pole.

307. Find the volume of the solid obtained by rotating about the x-axis the part of the parabola $y = ax - x^2$ which lies above the x-axis.

308. Find the volume of the solid obtained by rotating the ellipse $x^2/a^2 + y^2/b^2 = 1$ about the x-axis.

309. Find the volume generated by rotating about the x-axis the arc of the curve $y = \sin^2 x$ lying between $x = 0$ and $x = \pi$.

310. Find the volume generated by rotating the figure bounded by the x-axis, the semi-cubical parabola $y^2 = x^3$ and the ordinate $x = 1$ (a) about the x-axis, (b) about the y-axis.

311. Find the volume generated by rotating about the y-axis the area bounded by the parabola $y^2 = 4ax$ and the line $x = a$.

312. Find the volume generated by rotating about the x-axis the figure bounded by the curves $y = x^2$ and $y = \sqrt{x}$.

313. Find the volume generated by rotating, about the x-axis, the closed loop of the curve $(x - 4a)y^2 = ax(x - 3a)$ $(a > 0)$.

314. Find the volume traced out by rotating the curve $y^2 = x^3/(2a - x)$ about its asymptote $x = 2a$.

315. Find the volume of a paraboloid of revolution of radius of base R and height H.

316. A plane area bounded by a segment of a parabola of base $2a$ and height h is rotated about its base. Determine the volume of the solid so obtained (Cavalieri's "lemon").

317. Show that the volume cut off by the plane $x = 2a$ from the solid obtained by rotating the hyperbola $x^2 - y^2 = a^2$ about the x-axis is equal to that of a sphere of radius a.

318. Find the volume of the solid obtained by revolving the figure bounded by one arc of the cycloid $x = a(t - \sin t)$, $y = a(1 - \cos t)$ and the x-axis about (a) the x-axis, (b) the y-axis, (c) the axis of symmetry of the figure.

319. Find the volume of the solid generated by revolving the astroid $x = a\cos^3 t$, $y = b\sin^3 t$ about the y-axis.

320. Find the volume of the solid obtained by revolving the curve $r = a\cos^2\theta$ about the polar axis.

321. Find the volume of the solid obtained by revolving the curve $r = a(1 + \cos\theta)$ about the polar axis.

322. Find the volume of the solid cut off from the hyperboloid $x^2/a^2 + y^2/b^2 - z^2/c^2 = 1$ by the planes $z = 0$, $z = h$.

323. Find the volume of the ellipsoid $x^2/a^2 + y^2/b^2 + z^2/c^2 = 1$.

324. Find the area of the surface of the "spindle" which is formed by rotating one arc of the curve $y = \sin x$ about the x-axis.

325. Find the area of the surface formed by rotating about the x-axis the arc of the curve $y = \tan x$ lying between $x = 0$ and $x = \frac{1}{4}\pi$.

326. Find the area of the surface obtained by rotating the curve $y = e^{-x}$ $(0 \leqslant x \leqslant \infty)$ about the x-axis.

327. Find the area of the surface generated by the rotation about the x-axis of the arc of the catenary $y = a \cos h\,(x/a)$ lying between $x = 0$ and $x = a$.

328. Find the area of the surface obtained by rotating the astroid $x^{2/3} + y^{2/3} = a^{2/3}$ about the y-axis.

329. Find the area of the surface obtaining by rotating about the x-axis the part of the curve $x = \frac{1}{4}y^2 - \frac{1}{2}\log y$ which lies between $y = 1$ and $y = e$.

330. Find the surface area of the torus obtained by rotating the circle $x^2 + (y - b)^2 = a^2$ $(b > a)$ about Ox.

331. Find the surface areas of the solids obtained by rotating the ellipse $x^2/a^2 + y^2/b^2 = 1$ about (1) the x-axis; (2) the y-axis $(a > b)$.

332. Find the areas of the surfaces generated by rotating one arc of the cycloid $x = a\,(t - \sin t)$, $y = a\,(1 - \cos t)$ about (a) the x-axis; (b) the y-axis; (c) the tangent to the cycloid at its highest point.

333. Find the area of the surface obtained by rotating about the x-axis the cardioid with parametric equations $x = a\,(2 \cos t -$ $- \cos 2t)$, $y = a\,(2 \sin t - \sin 2t)$.

334. Find the statical moments with respect to the coordinate axes of the segment of the line $x/a + y/b = 1$ cut off by the coordinate axes.

335. Find the statical moments of a rectangle of sides a and b with respect to its sides.

336. Find the statical moments with respect to the coordinate axes and the coordinates of the centroid of the area bounded by $x + y = a$, $x = 0$, $y = 0$.

337. Find the statical moments with respect to the coordinate axes and the coordinates of the centroid of the area in the first quadrant bounded by the astroid.

338. Find the statical moment of the circle $r = 2a \sin \theta$ with respect to the polar axis.

339. Find the centroid of the arc of a circle of radius a subtending an angle 2α at the centre of the circle.

340. Find the coordinates of the centroid of one arc of the cycloid $x = a(t - \sin t)$, $y = a(1 - \cos t)$.

341. Find the coordinates of the area in the first quadrant bounded by the coordinate axes and the ellipse $x^2/a^2 + y^2/b^2 = 1$.

342. Find the coordinates of the centroid of the area bounded by the parabolas $y = x^2$, $y = \sqrt{x}$.

343. Find the coordinates of the centroid of the area bounded by one arc of the cycloid $x = a(t - \sin t)$, $y = a(1 - \cos t)$.

344. Find the centroid of a hemispherical shell of radius a, taking the centre of the sphere to be the origin of coordinates and the plane face to be the plane $x\,0\,y$.

345. Find the centroid of a homogeneous right circular cone of height h, taking the origin of coordinates at the apex and the z-axis along the axis of symmetry.

346. Find the centroid of a homogeneous hemisphere of radius a, taking the same coordinate system as in **344.**

347. Find the moment of inertia of a circular hoop of radius a about a diameter.

348. Find the moments of inertia of a rectangle of sides a, b about its sides.

349. Find the moment of inertia of the plane area bounded by a parabolic segment with base $2b$ and height h about its axis of symmetry.

350. Find the moments of inertia of the area bounded by the ellipse $x^2/a^2 + y^2/b^2 = 1$ about its principal axes.

351. Find the moment of inertia about an axis through its centre and normal to a plane of a circular ring of radii r_1, r_2 $(r_2 > r_1)$.

352. Find the moment of inertia of a homogeneous right circular cone of radius r and height h about its axis.

353. Find the moment of inertia of a solid sphere of radius a and mass m about a diameter.

354. Find the volume and surface area of the torus obtained by rotating a circle of radius a about an axis in the same plane and distant b $(> a)$ from the centre of the circle.

355. (a) Find the centroid of a semicircle using Guldin's theorem.
(b) Using Guldin's theorem, prove that the centroid of a triangle is at distance $\frac{1}{3}h$ from the base, where h is the height of the triangle.

CHAPTER IV

SERIES. APPLICATIONS TO APPROXIMATE EVALUATIONS

§ 12. Basic theory of infinite series

118. Infinite series. Let an infinite sequence of numbers be given:

$$u_1, u_2, u_3, \ldots, u_n, \ldots \tag{1}$$

By taking the sum of the first n terms

$$s_n = u_1 + u_2 + \ldots + u_n\,, \tag{2}$$

we can get another infinite sequence of numbers

$$s_1, s_2, \ldots, s_n, \ldots$$

If s_n tends to a (finite) limit on indefinite increase of n :

$$s = \lim_{n\to\infty} s_n\,,$$

we say that the infinite series :

$$u_1 + u_2 + \ldots + u_n + \ldots \tag{3}$$

converges and has the sum s, and we write :

$$s = u_1 + u_2 + \ldots + u_n + \ldots \tag{4}$$

On the other hand, if s_n does not tend to a limit, we say that the infinite series (3) *diverges.*

In other words *the infinite series* (3) *is said to converge, if the sum of its first n terms tends to a limit on indefinite increase of n. We call this limit the sum of the series.*

We can only talk about the sum of an infinite series when it is convergent, in which case the sum s_n of the first n terms gives an approx-

imate expression for the sum s of the series. The error r_n involved in this approximation, i.e. the difference

$$r_n = s - s_n \, ,$$

is called the *remainder* of the series.

The remainder r_n is evidently itself the sum of an infinite series, obtained from the given series (3) by neglecting its first n terms:

$$r_n = u_{n+1} + u_{n+2} + \ldots + u_{n+p} + \ldots$$

This remainder cannot be found accurately in the majority of cases, so that it is important to know the *approximate error* due to this remainder.

The simplest example of an infinite series is the geometrical progression:

$$a + aq + aq^2 + \ldots + aq^{n-1} + \ldots \quad (a \neq 0) \tag{5}$$

We consider separately the cases:

$$|\, q \,| < 1, \ |\, q \,| > 1, \ q = 1, \ q = -1.$$

We know [27] that a geometrical progression has a finite sum $s = a/(1-q)$ for $|q| < 1$, so that the series is convergent; here, in fact:

$$s_n = a + aq + \ldots + aq^{n-1} = \frac{a - aq^n}{1-q},$$

$$s - s_n = \frac{a}{1-q} - \frac{a - aq^n}{1-q} = \frac{aq^n}{1-q},$$

and $s - s_n \to 0$ as $n \to \infty$, since $q_n \to 0$ for $|\, q \,| < 1$ [26]. Obviously, $s_n \to \infty$ for $n \to \infty$ with $|\, q \,| > 1$, since now $q^n \to \infty$ [29]. We have $s_n = an$ for $q = 1$, and evidently $s_n \to \infty$ again, so that a geometrical progression is divergent for $|\, q \,| > 1$ and $q = 1$. We get the series, with $q = -1$:

$$a - a + a - a + \ldots$$

The sum s_n of its first n terms is zero for n even, and a for n odd, i.e. s_n does not tend to a limit, and the series is divergent; though contrary to the previous case, the sum remains bounded for all n, since it only takes the values 0 and a.

If the absolute value of s_n, the sum of the first n terms of series (3), *tends to infinity on indefinite increase of n, series* (1) *is said to be strictly divergent.* In future, we shall simply speak of a "divergent series" for brevity, when a strictly divergent series is meant.

119. Basic properties of infinite series. Convergent infinite series have certain properties that allow for operations being performed on them, as on finite sums.

I. *If the series*

$$u_1 + u_2 + \dots + u_n + \dots$$

has the sum s, *the series*

$$au_1 + au_2 + \dots + au_n + \dots, \tag{6}$$

obtained by multiplying all the terms of the first series by the same factor a, *has the sum* as, because the sum σ_n of the first n terms of (6) is

$$\sigma_n = au_1 + au_2 + \dots + au_n = as_n\,,$$

and hence

$$\lim_{n\to\infty} \sigma_n = \lim_{n\to\infty} as_n = a \lim_{n\to\infty} s_n = as.$$

II. *Convergent series can be added and subtracted term by term, i.e. if*

$$u_1 + u_2 + \dots + u_n + \dots = s\,,$$

$$v_1 + v_2 + \dots + v_n + \dots = \sigma\,,$$

the series

$$(u_1 \pm v_1) + (u_2 \pm v_2) + \dots + (u_n \pm v_n) + \dots, \tag{7}$$

is also convergent, and its sum is $(s \pm \sigma)$, since the sum of the first n terms of (7) is

$$(u_1 \pm v_1) + (u_2 \pm v_2) + \dots + (u_n \pm v_n) = s_n \pm \sigma_n.$$

Other properties of sums, for instance the independence of the sum on the order of the terms, the rule for cross-multiplying two sums, etc., will be considered in regard to infinite series in § 14 below. We remark for the present that these properties do not hold for every series. The associative law is evidently valid for any convergent series, i.e. we can regroup the terms of the series as desired. This amounts to taking only part of the s_n instead of all the s_n ($n = 1, 2, 3, \dots$), which does not alter the limit s.

III. *The convergence or divergence of a series is unaltered by removing or adding a finite number of terms at the beginning.* We shall take the two series:

$$u_1 + u_2 + u_3 + u_4 + \dots$$

$$u_3 + u_4 + u_5 + u_6 + \dots$$

The second is derived from the first by removing the first two terms. Let the sum of the first n terms of the first series be denoted by s_n,

and of the second series by σ_n. Evidently:

$$\sigma_{n-2} = s_n - (u_1 + u_2), \quad s_n = \sigma_{n-2} + (u_1 + u_2),$$

whilst subscript $(n - 2) \to \infty$ as $n \to \infty$. Hence, if s_n has a limit, σ_{n-2} also has a limit, and conversely. These limits, s and σ, i.e. the sums of the above two series, will of course be different, and in fact: $\sigma = s - (u_1 + u_2)$.

IV. *The general term u_n of a convergent series tends to zero on indefinite increase of n :*

$$\lim u_n = 0\,, \tag{8}$$

since obviously,

$$u_n = s_n - s_{n-1}\,,$$

and if the series converges and has the sum s,

$$\lim s_{n-1} = \lim s_n = s\,,$$

so that

$$\lim u_n = \lim s_n - \lim s_{n-1} = s - s = 0.$$

Condition (8) is thus necessary for the convergence of a series, but it is not sufficient: a series as a whole can diverge, whilst its general terms tends to zero.

Example. The harmonic series is:

$$1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \ldots = \sum_{n=1}^{\infty} \frac{1}{n}\,. \tag{9}$$

We have here:

$$u_n = \frac{1}{n} \to 0 \text{ for } n \to \infty.$$

It can easily be shown, however, that the sum of the first n terms of (9) increases indefinitely. We take the terms in groups of 1, 2, 4, 8, ... terms, starting with the second:

$$1 + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \ldots + \frac{1}{8}\right) + \left(\frac{1}{9} + \ldots + \frac{1}{16}\right) + \ldots$$

so that the kth group contains 2^{k-1} terms. If we replace all the terms of each group by the last, this being the smallest of the group, the resulting series:

$$1 + \frac{1}{2} + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 4 + \ldots = 1 + \frac{1}{2} + \frac{1}{2} + \ldots, \tag{10}$$

evidently tends to $(+\infty)$, the sum of its first n terms being $[1 + {}^1/_2\,(n - 1)]$. We can take any desired number n of groups by taking a sufficiently large number of terms of (9), and the sum of the terms of (9) is greater than $[1 + {}^1/_2\,(n - 1)]$, whence it follows that $s_n \to +\infty$ for the series (9).

120. Series with positive terms. Tests for convergence. Series with positive (non-negative) terms have a special importance; here, all the terms

$$u_1, u_2, u_3, \ldots, u_n, \ldots \geqslant 0.$$

We establish a number of tests for convergence and divergence of such series.

1. *A series with positive terms can only be convergent, or else strictly divergent, i.e.*

$$s_n \to s \text{ or } s_n \to +\infty.$$

A necessary and sufficient condition for the convergence of a series with positive terms is that the sum s_n of its first n terms remains less than a given constant A for any n.

The sum s_n of the series cannot decrease as n increases, since new positive (non-negative) terms are then added; all our assertions now follow from the properties of increasing variables worked out in [30].

It is often useful, when discussing the convergence or divergence of a series with positive terms, to compare it with other, simpler series, and especially with a geometrical progression.

We therefore establish the following test:

2. *If each term of a series with positive terms*

$$u_1 + u_2 + \ldots + u_n + \ldots, \tag{11}$$

as from some given term, does not exceed the corresponding term of a convergent series

$$v_1 + v_2 + \ldots + v_n + \ldots, \tag{12}$$

the given series is convergent.

If, on the contrary, each term of (11), *as from a given n, is not less than the corresponding term of a divergent series* (12), *the given series is also divergent.*

We first take

$$u_n \leqslant v_n, \tag{13}$$

with series (12) convergent. We can suppose without loss of generality that this inequality is true for all n, since the first terms to which it does not apply can be neglected if necessary (property III [119]). Let the sum of the first n terms of (11) be denoted by s_n, and the corresponding sum for (12) by σ_n; then we have by (13):

$$s_n \leqslant \sigma_n.$$

But (12) converges by hypothesis, and denoting its sum by σ, we have:

$$\sigma_n \leqslant \sigma\,,$$

so that we also have:

$$s_n \leqslant \sigma\,,$$

whence the convergence of (11) follows, by 1.

We now take

$$u_n \geqslant v_n. \tag{14}$$

We evidently have

$$s_n \geqslant \sigma_n\,; \tag{15}$$

but series (12) is now divergent, and the sum σ_n of its first n terms can be made greater than any previously assigned large number; by (15), s_n has the same property, i.e. series (11) is also divergent.

Remark. The convergence (divergence) of series (12) implies the convergence (divergence) of the series

$$kv_1 + kv_2 + kv_3 + \ldots + kv_n + \ldots,$$

where k is any positive constant number.

The convergence of the series $\Sigma\, kv_n$ follows from the convergence of $\Sigma\, v_n$ by I [119]. Conversely, if Σv_n is divergent, $\Sigma\, kv_n$ must be divergent, since, supposing it were convergent, and multiplying its terms by $1/k$, we should have the convergence of $\Sigma\, v_n$ by I [119]. It follows from what has been said that:

Series (11) *is convergent, if*

$$u_n \leqslant kv_n\,, \tag{16}$$

where the series $\Sigma\, v_n$ is convergent and k is any positive number ; (11) *is divergent, if*

$$u_n \geqslant kv_n\,, \tag{17}$$

where $\Sigma\, v_n$ is divergent.

By comparing a given series with a geometrical progression, we obtain two fundamental tests for the convergence of series with positive terms.

121. Cauchy's and d'Alembert's tests.

3. Cauchy's test. *If the general term of the series* (11) *with positive terms :*

$$u_1 + u_2 + \ldots + u_n + \ldots,$$

satisfies, as from some given n:

$$\sqrt[n]{u_n} \leqslant q < 1, \tag{18}$$

where q does not depend on n, the series is convergent.

On the other hand, if we have, as from a given n,

$$\sqrt[n]{u_n} \geqslant 1, \tag{19}$$

series (11) *is divergent.*

We can assume that inequality (18) or (19) is satisfied for all n without loss of generality (property III [119]). If (18) is satisfied,

$$u_n \leqslant q^n,$$

i.e. the general term of the given series does not exceed the corresponding term of an infinite diminishing geometrical progression, whence, by test 2., the series is convergent. Whereas we have in the case of (19):

$$u_n \geqslant 1,$$

and series (11) has a general term which does not tend to zero (it is greater than unity), so that the series cannot be convergent (property IV [119]).

4. d'Alembert's test. *If the ratio u_n/u_{n-1} of two successive terms of a series satisfies the inequality, as from a given n :*

$$\frac{u_n}{u_{n-1}} \leqslant q < 1, \tag{20}$$

where q does not depend on n, series (11) *is convergent.*

On the other hand, if we have, as from a given n :

$$\frac{u_n}{u_{n-1}} \geqslant 1, \tag{21}$$

the given series is divergent.

We can make the same sort of assumption as before, that inequality (20) or (21) is satisfied for all n; we have in the case of (20):

$$u_n \leqslant u_{n-1}\, q,\ u_{n-1} \leqslant u_{n-2}\, q,\ \ldots,\ u_2 \leqslant u_1\, q,$$

whence, cross-multiplying term by term and cancelling the common factors,

$$u_n \leqslant u_1\, q^{n-1},$$

i.e. the terms of the series are less than the terms of the diminishing geometrical progression:

$$u_1 + u_1 q + u_1 q^2 + \ldots + u_1 q^{n-1} + \ldots \quad (0 < q < 1),$$

and (11) therefore converges, by test 2. In the case of (21):

$$u_1 \leqslant u_2 \leqslant u_3 \leqslant \ldots \leqslant u_{n-1} \leqslant u_n \ldots,$$

i.e. the terms of the series do not decrease, so that u_n does not tend to zero for $n \to \infty$, and the series cannot be convergent (property IV [119]).

Corollary. *If*

$$\sqrt[n]{u_n} \quad \text{or} \quad \frac{u_n}{u_{n-1}} \tag{22}$$

tends to a finite limit r on indefinite increase of n, the series

$$u_1 + u_2 + \ldots + u_n + \ldots$$

is convergent for $r < 1$, and divergent for $r > 1$.

Suppose first $r < 1$. We take a sufficiently small number ε such that also

$$r + \varepsilon < 1.$$

For large n, $\sqrt[n]{u_n}$ or u_n/u_{n-1} will differ from the limit r by not more than ε, i.e. we shall have, starting from some sufficiently large n:

$$r - \varepsilon \leqslant \sqrt[n]{u_n} \leqslant r + \varepsilon < 1 \tag{23_1}$$

or

$$r - \varepsilon \leqslant \frac{u_n}{u_{n-1}} \leqslant r + \varepsilon < 1\,. \tag{23_2}$$

Applying Cauchy's or d'Alembert's test with $q = r + \varepsilon < 1$, we can infer at once the convergence of the given series from (23_1) or (23_2).

The proof of divergence is similar, in the case when $r > 1$. The series is evidently also divergent if one of expressions (22) tends to $(+\infty)$.

Examples. **1.** Consider the series

$$1 + \frac{x}{1} + \frac{x^2}{1.2} + \ldots + \frac{x^n}{1.2.3 \ldots n} + \ldots = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{24}$$

Applying d'Alembert's test:

$$u_{n+1} = \frac{x^n}{n!}, \quad u_n = \frac{x^{n-1}}{(n-1)!}, \quad \frac{u_{n+1}}{u_n} = \frac{x}{n} \to 0 \text{ as } n \to \infty\,,$$

so that the given series converges for all finite (positive) x.

2. Consider the series

$$\sum_{n=1}^{\infty} \frac{x^n}{n}\,. \tag{25}$$

Here we have:

$$u_n = \frac{x^n}{n}, \quad u_{n-1} = \frac{x^{n-1}}{n-1}, \quad \frac{u_n}{u_{n-1}} = \frac{n-1}{n} x \to x,$$

so that, by d'Alembert's test, the given series converges for $0 \leqslant x < 1$ and diverges for $x > 1$.

3. Consider the series

$$\sum_{n=1}^{\infty} r^n \sin^2 n\,a. \tag{26}$$

We have, on applying Cauchy's test:

$$u_n = r^n \sin^2 n\,a, \quad \sqrt[n]{u_n} = r\sqrt[n]{\sin^2 n\,a} \leqslant r,$$

so that the series is convergent for $r < 1$.

d'Alembert's test gives no result in this case, since

$$\frac{u_n}{u_{n-1}} = r\left[\frac{\sin n\,a}{\sin (n-1)\,a}\right]^2$$

neither tends to a limit, nor remains always < 1 or $\geqslant 1$.

It can be shown that, in general, Cauchy's test is stronger than d'Alembert's test, i.e. Cauchy's test can always be applied when d'Alembert's can, and often when it cannot. On the other hand, d'Alembert's test is easier to use, as becomes evident from an inspection of the first two examples worked out above.

We remark further that there are cases when both Cauchy's and d'Alembert's tests fail, as, for instance, whenever

$$\sqrt[n]{u_n} \quad \text{and} \quad \frac{u_n}{u_{n-1}} \to 1,$$

i.e. when $r = 1$. We are concerned here with a *doubtful* case, when the question of convergence or divergence has to be decided by some other means.

For example, in the case of the harmonic series

$$\sum_{n=1}^{\infty} \frac{1}{n},$$

which we have seen to be divergent in [119], we have:

$$\frac{u_n}{u_{n-1}} = \frac{n-1}{n} \to 1, \quad \sqrt[n]{u_n} = \sqrt[n]{\frac{1}{n}} = e^{\frac{1}{n}\log\frac{1}{n}} \to 1,\dagger$$

† It must be noted in the above working that, putting $x = 1/n$, $x \to 0$ and $(1/n) \log 1/n = x \log x \to 0$ [66]. Hence, on taking the logarithm of the expression $\sqrt[n]{1/n}$, we see that it tends to unity.

and therefore the question of the convergence or divergence of the harmonic series cannot be decided by means of Cauchy's or d'Alembert's test.

We show later that the series

$$\sum_{n=1}^{\infty} \frac{1}{n^2} = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \ldots$$

is *convergent.*

But we again have in this case:

$$\frac{u_n}{u_{n-1}} = \left(\frac{n-1}{n}\right)^2 \to 1\,, \quad \sqrt[n]{u_n} = \sqrt[n]{\frac{1}{n^2}} = \left(\sqrt[n]{\frac{1}{n}}\right)^2 \to 1\,,$$

i.e. another doubtful case, if Cauchy's or d'Alembert's test is applied.

122. Cauchy's integral test for convergence. We assume that the terms of the series:

$$u_1 + u_2 + u_3 + \ldots + u_n + \ldots \tag{27}$$

are positive and non-increasing, i. e.

$$u_1 \geqslant u_2 \geqslant \ldots \geqslant u_n \geqslant u_{n+1} \geqslant \ldots \geqslant 0\,. \tag{28}$$

We represent the terms graphically by plotting n as the independent variable, which as yet only takes integral values, along the axis of abscissae, and the corresponding values of u_n on the axis of ordinates (Fig. 155). A continuous function $y=f(x)$ can always be found, such that it takes precisely the values u_n for integral values of $x = n$; in fact, it is only necessary to draw a continuous curve through all the points plotted; we shall assume that the function thus obtained is non-increasing.

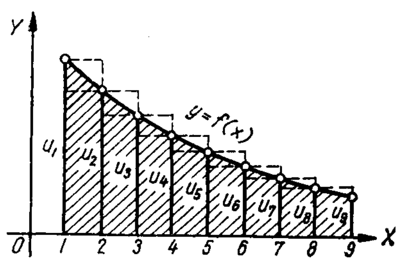

FIG. 155

With this graphical representation, the sum of the first n terms of the series,

$$s_n = u_1 + u_2 + \ldots + u_n\,,$$

is represented by the sum of the areas of the exterior rectangles in which is contained the area of the figure bounded by the curve $y = f(x)$, axis OX, and ordinates $x = 1$ and $x = n + 1$, so that

$$s_n \geqslant \int_1^{n+1} f(x)\,dx\,. \tag{29}$$

On the other hand, this latter figure contains within itself all the interior rectangles, the sum of the areas of which is:

$$u_2 + u_3 + u_4 + \ldots + u_{n+1} = s_{n+1} - u_1 , \tag{30}$$

so that

$$s_{n+1} - u_1 \leqslant \int_1^{n+1} f(x)\,dx . \tag{31}$$

These inequalities lead us to the following test.

5. Cauchy's integral test. *The series* (27),

$$u_1 + u_2 + \ldots + u_n + \ldots , \quad u_n = f(n) ,$$

with positive terms that do not increase with increasing n, converges or strictly diverges, according as the integral

$$I = \int_1^{\infty} f(x)\,dx \tag{32}$$

is finite or infinite.

We recall that the $f(x)$ here must decrease with increasing x.

Firstly, let integral I be finite, i.e. the curve $y = f(x)$ has a finite area [98]. It follows from the fact that $f(x)$ is positive, that

$$\int_1^{n+1} f(x)\,dx < \int_1^{\infty} f(x)\,dx ,$$

and hence, by (31):

$$s_n < s_{n+1} \leqslant u_1 + I ,$$

i.e. the sum s_n is bounded for all n, and (27) is convergent, on the basis of test 1 [120].

We now take $I = \infty$, i.e. the integral

$$\int_1^{n+1} f(x) dx$$

can be made greater than any previously assigned number N by increasing n. By (29), s_n can also be made greater than N, i.e. (27) is strictly divergent.

It can similarly be shown that *the remainder of series* (27) *does not exceed the integral*

$$\int_n^{\infty} f(x)\,dx .$$

Remark. The integral I in Cauchy's test can be replaced by the integral

$$\int_a^\infty f(x)\,dx$$

where a is any positive number, greater than unity.

In fact, if the curve $y = f(x)$ has a finite area on measuring from the ordinate $x = 1$, its area will be finite on measuring from any ordinate $x = a$, and *vice versa*. If $I = \infty$, we speak alternatively of integral (32) diverging.

Examples. **1.** Consider the harmonic series:

$$\sum_{n=1}^\infty \frac{1}{n}.$$

Here we have:

$$f(n) = \frac{1}{n},$$

so that we can put

$$f(x) = \frac{1}{x}:$$

then

$$I = \int_1^\infty \frac{dx}{x} = \log x \Big|_1^\infty$$

and the integral diverges, since $\log x \to +\infty$ as $x \to +\infty$; thus the series diverges, as we have already seen.

2. Consider the more general series

$$\sum_{n=1}^\infty \frac{1}{n^p}, \tag{33}$$

where p is any number greater than zero (the series obviously diverges for $p \leqslant 0$). Here we have:

$$f(n) = \frac{1}{n^p}, \quad f(x) = \frac{1}{x^p}, \quad I = \int_1^\infty \frac{dx}{x^p} = \begin{cases} \dfrac{1}{1-p} x^{1-p} \Big|_1^\infty \ \textit{if}\ p \neq 1, \\ \log x \Big|_1^\infty \ \textit{if}\ p = 1. \end{cases}$$

Hence it is clear that the integral diverges, if $p \leqslant 1$, and is convergent to $1/(p-1)$, if $p > 1$. We have in the latter case the exponent $1 - p < 0$, $x^{1-p} = 1/x^{p-1} \to 0$ as $x \to +\infty$, and thus

$$\frac{1}{1-p} x^{1-p} \Big|_1^\infty = 0 - \frac{1}{1-p} = \frac{1}{p-1}.$$

Hence, by Cauchy's test, series (33) is *convergent for $p > 1$, and divergent for $p \leqslant 1$.*

123. Alternating series. We now turn to series with either positive or negative terms, and we first take the *alternating* series, the terms of which are alternately positive and negative. Such series are more conveniently written in a new form:

$$u_1 - u_2 + u_3 - u_4 + \ldots \pm u_n \mp u_{n+1} \ldots, \tag{34}$$

where the numbers

$$u_1, u_2, u_3, \ldots u_n, \ldots$$

are reckoned positive.†

The following proposition can be proved regarding alternating series:

A sufficient condition for the convergence of an alternating series is that the absolute value of its terms decrease and tend to zero with increasing n. The absolute value of the remainder of the series does not exceed the absolute value of the first of the neglected terms.

We first take the sum of an even number of terms of the series:

$$s_{2n} = u_1 - u_2 + u_3 - \ldots + u_{2n-1} - u_{2n}.$$

Since, by hypothesis, the absolute values of the terms decrease (or it is better to say, do not increase) with increasing n, in general:

$$u_k \geqslant u_{k+1} \quad \text{and} \quad u_{2n+1} - u_{2n+2} \geqslant 0\,,$$

and hence

$$s_{2n+2} = s_{2n} + (u_{2n+1} - u_{2n+2}) \geqslant s_{2n}\,,$$

i.e. the variable s_{2n} is non-decreasing. We have on the other hand:

$$s_{2n} = u_1 - (u_2 - u_3) - (u_4 - u_5) - \ldots - (u_{2n-2} - u_{2n-1}) - u_{2n} \leqslant u_1\,,$$

since all the differences in brackets are non-negative; i.e. the variable s_{2n} remains bounded for all n. Hence it follows that, on indefinite increase of n, s_{2n} tends to a finite limit [30], which we denote by s:

$$\lim_{n\to\infty} s_{2n} = s\,.$$

We have further:

$$s_{2n+1} = s_{2n} + u_{2n+1} \to s \quad as \quad n \to \infty\,,$$

since $u_{2n+1} \to 0$ by hypothesis.

† We assume here that the first term of the series is positive; if it is negative, the series is written as $-u_1 + u_2 - u_3 + \ldots$

We thus see that the sum of both an even and odd number of terms of series (34) tends to the same limit s, i.e. (34) is convergent and has the sum s.

The remainder r_n of the series has still to be found. We have:

$$r_n = \pm u_{n+1} \mp u_{n+2} \pm u_{n+3} \dots,$$

where either the upper or lower signs have to be taken simultaneously. Alternatively,

$$r_n = \pm (u_{n+1} - u_{n+2} + u_{n+3} - \dots)$$

whence, by the same argument as before:

$$|r_n| = (u_{n+1} - u_{n+2}) + (u_{n+3} - u_{n+4}) + \dots =$$
$$= u_{n+1} - (u_{n+2} - u_{n+3}) - \dots \leqslant u_{n+1},$$

which it was required to prove.

It follows from the formula:

$$r_n = \pm \left[(u_{n+1} - u_{n+2}) + (u_{n+3} - u_{n+4}) + \dots\right],$$

where the square brackets contain a non-negative quantity, that the sign of r_n is the same as that to be taken in front of the square brackets, i.e. is the same as the sign of $+u_{n+1}$. Thus, *with the hypotheses laid down in the theorem, the sign of the remainder of an alternating series is the same as that of the first neglected term.*

Example. The absolute values of the terms of the alternating series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \dots$$

decrease indefinitely as $n \to \infty$, so that the series is convergent. We see later that its sum is log 2. The series is not suitable, however, for the actual computation of log 2, since we should have to take 10,000 terms to obtain a remainder less than 0.0001:

$$|r_n| < \frac{1}{n+1} \leqslant 0.0001; \quad n \geqslant 10{,}000.$$

This series, though convergent, *converges very slowly*; in order to deal with this type of series in practice, a preliminary transformation is needed, so as to pass from the slowly convergent to a rapidly convergent series, or, as we say, to improve the convergence.

124. Absolutely convergent series. As regards other series with terms of any sign, we shall restrict our attention to those that are absolutely convergent.

The series

$$u_1 + u_2 + u_3 + \dots + u_n + \dots \tag{35}$$

is convergent, if the series consisting of the absolute values of its terms is convergent, i.e. *if the series*

$$|u_1| + |u_2| + |u_3| + |u_4| + \ldots + |u_n| + \ldots \tag{36}$$

is convergent.

A series of this sort is called absolutely convergent.

Let us suppose that (36) is convergent; we put

$$v_n = \frac{1}{2}(|u_n| + u_n), \qquad w_n = \frac{1}{2}(|u_n| - u_n).$$

Both v_n and w_n are non-negative numbers, since obviously

$$v_n = \begin{cases} u_n, & \text{if} \quad u_n \geqslant 0, \\ 0 & ,, \quad u_n \leqslant 0, \end{cases} \qquad w_n = \begin{cases} 0, & \text{if} \quad u_n \geqslant 0. \\ |u_n|, & ,, \quad u_n \leqslant 0. \end{cases}$$

On the other hand, neither v_n nor w_n exceeds $|u_n|$, which is the general term of the convergent series (36), and hence, by the second test for the convergence of series with positive terms [120], both series

$$\sum_{n=1}^{\infty} v_n, \quad \sum_{n=1}^{\infty} w_n$$

will be convergent.

Since we have:

$$u_n = v_n - w_n,$$

the series

$$\sum_{n=1}^{\infty} u_n = \sum_{n=1}^{\infty} (v_n - w_n) = \sum_{n=1}^{\infty} v_n - \sum_{n=1}^{\infty} w_n,$$

got by term by term subtraction of $\sum_{n=1}^{\infty} w_n$ from $\sum_{n=1}^{\infty} v_n$, is convergent [119].

Convergent series with positive terms represent a particular case of absolutely convergent series. Convergence tests for these latter are obtained directly from the tests for series with positive terms.

Convergence tests 1—5, *deduced in* [120, 121, 122] *for series with positive terms, can be used for series with terms of any sign, provided only that we agree to replace* u_n *everywhere by* $|u_n|$. *With this proviso, divergence tests* 3 *and* 4 *remain valid, together with their corollary* [121].

In particular, in formulating Cauchy's and d'Alembert's tests, we have to replace:

$$\sqrt[n]{u_n} \quad \text{and} \quad \frac{u_n}{u_{n-1}} \quad \text{by} \quad \sqrt[n]{|u_n|} \quad \text{and} \quad \left|\frac{u_n}{u_{n-1}}\right|.$$

If, for instance, $|u_n/u_{n-1}| < q < 1$, i.e. $|u_n|/|u_{n-1}| < q < 1$, series (36) with positive terms is convergent by d'Alembert's test [121], and series (35) is therefore absolutely convergent. On the other hand, if $|u_n/u_{n-1}| \geqslant 1$, i.e. $|u_n| \geqslant |u_{n-1}|$, the term u_n cannot decrease in absolute value with increasing n, and therefore cannot tend to zero, and series (35) is divergent. Hence it follows, as in the corollary of [121], that if $|u_n/u_{n-1}| \to r < 1$, series (35) is absolutely convergent; whilst if $|u_n/u_{n-1}| \to r > 1$, series (35) is divergent.

Remark. We note further that, if the absolute values of the terms of a series (35) *are not greater than the positive numbers* a_n, $|u_n| \leqslant a_n$, *the series of these numbers being convergent,* i.e. $a_1 + a_2 + \ldots + a_n + \ldots$ *is convergent, series* (36) *is then certainly convergent* [120], i.e. *series* (35) *is absolutely convergent.*

Examples. **1.** The series (Example [121])

$$\sum_{n=1}^{\infty} \frac{x^n}{n!}$$

is absolutely convergent for all finite x, either positive or negative, since

$$\left|\frac{u_{n+1}}{u_n}\right| = \frac{|x|}{n} \to 0$$

for all finite x.

2. The series

$$\sum_{n=1}^{\infty} \frac{x^n}{n}$$

is absolutely convergent for $|x| < 1$, and divergent for $|x| > 1$, since

$$\left|\frac{u_n}{u_{n-1}}\right| = \frac{n-1}{n}|x| \to |x|.$$

3. The series

$$\sum_{n=1}^{\infty} r^n \sin na$$

is absolutely convergent for $|r| < 1$, since then:

$$\sqrt[n]{|u_n|} = \sqrt[n]{|r^n|\,|\sin na|} \leqslant \sqrt[n]{|r|^n} = |r| < 1.$$

It must be noted that by no means every convergent series is also absolutely convergent, i.e. remains convergent after replacing each term of the series by its absolute value. For instance, the alternating series

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

is convergent, as we have seen; whereas on replacing each term by its absolute value, we get the divergent harmonic series:

$$1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$$

Absolutely convergent series have a number of interesting properties, discussed in the small print section § 14. For example, they alone have a property of finite sums, namely independence of the sum on the order of the terms.

125. General test for convergence. We conclude the present section with some remarks on the necessary and sufficient condition for convergence of the series

$$u_1 + u_2 + u_3 + \dots + u_n + \dots$$

Its convergence is, by definition, equivalent to the existence of a limit for the sequence:

$$s_1, s_2, \dots, s_n, \dots,$$

where s_n is the sum of the first n terms of the series. But we have Cauchy's necessary and sufficient condition for the existence of this limit [31]:

For any given positive ε there exists an N, such that

$$| s_m - s_n | < \varepsilon$$

for all m and $n > N$. We take $m > n$ for clarity, and let $m = n + p$, where p is any positive integer. Noting that now

$$s_m - s_n = s_{n+p} - s_n = (u_1 + u_2 + \dots + u_n + u_{n+1} + \dots + u_{n+p}) -$$
$$- (u_1 + u_2 + \dots + u_n) = u_{n+1} + u_{n+2} \dots + u_{n+p},$$

we can put forward the following general test for the convergence of a series.

A necessary and sufficient condition for the convergence of the infinite series

$$u_1 + u_2 + u_3 + \dots + u_n + \dots$$

is that for any previously assigned positive ε *there exists an* N, *such that, for any* $n > N$ *and for any positive* p,

$$| u_{n+1} + u_{n+2} + \dots + u_{n+p} | < \varepsilon,$$

i.e. *starting with* u_{n+1}, *the absolute value of the sum of any number of subsequent terms of the series remains less than* ε *as soon as* $n > N$.

It must be remarked that this general test for the convergence of a series, though of great theoretical importance, is usually difficult to apply in practice.

§ 13. Taylor's formula and its applications

126. Taylor's formula. We take a polynomial of degree n:

$$f(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n ;$$

we give x an increment h and find the corresponding value of the function $f(x + h)$. This value can evidently be expressed in powers of h, by expanding the various powers of $(x + h)$ by means of Newton's binomial formula and re-arranging the result in powers of h. The coefficients of these powers of h will be polynomials in x:

$$f(x+h) = A_0(x) + hA_1(x) + h^2A_2(x) + \dots + h^kA_k(x) + \dots + h^nA_n(x), \tag{1}$$

and we have to determine:

$$A_0(x), A_1(x), \dots, A_n(x).$$

We first change the notation, writing a instead of x, and x instead of $(x + h)$, in the identity (1). We now have

$$h = x - a ,$$

and (1) becomes:

$$f(x) = A_0(a) + (x - a)A_1(a) + (x - a)^2A_2(a) + \dots + (x - a)^kA_k(a) + \dots + (x - a)^nA_n(a). \tag{2}$$

We find $A_0(a)$ by putting $x = a$ in this identity, which gives:

$$f(a) = A_0(a).$$

We find $A_1(a)$ by differentiating (2) with respect to x then putting $x = a$:

$$f'(x) = 1 \times A_1(a) + 2(x - a)A_2(a) + \dots + k(x - a)^{k-1}A_k(a) + \dots + n(x - a)^{n-1}A_n(a),$$

$$f'(a) = 1 \times A_1(a).$$

We differentiate a second time with respect to x, then put $x = a$, and find $A_2(a)$:

$$f''(x) = 2 \times 1 A_2(a) + \dots + k(k-1)(x-a)^{k-2} A_k(a) + \dots + \\ + n(n-1)(x-a)^{n-2} A_n(a).$$

$$f''(a) = 2 \times 1 A_2(a).$$

Continuing these operations, we differentiate k times with respect to x then put $x = a$, and get:

$$f^{(k)}(x) = k(k-1)\dots 2 \times 1 A_k(a) + \dots + \\ + n(n-1)\dots(n-k+1)(x-a)^{n-k} A_n(a),$$

$$f^{(k)}(a) = k!\, A_k(a).$$

We thus have:

$$A_0(a) = f(a), \quad A_1(a) = \frac{f'(a)}{1!}, \quad A_2(a) = \frac{f''(a)}{2!}, \dots,$$

$$A_k(a) = \frac{f^{(k)}(a)}{k!}, \dots, A_n(a) = \frac{f^{(n)}(a)}{n!},$$

so that (2) now reads:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \\ + \frac{f^{(k)}(a)}{k!}(x-a)^k + \dots + \frac{f^{(n)}(a)}{n!}(x-a)^n. \tag{3}$$

This formula is only true when $f(x)$ is a polynomial of degree not greater than n, in which case it gives an expression for the polynomial in powers of the difference $(x - a)$. Let $f(x)$ be any given function, allowing for differentiation up to and including the nth order. Let $R_n(x)$ be the error obtained by taking the right-hand side of (3) for $f(x)$, i.e. we have:

$$f(x) = f(a) + \frac{f'(a)}{1!}(x-a) + \frac{f''(a)}{2!}(x-a)^2 + \dots + \\ + \frac{f^{(n)}(a)}{n!}(x-a)^n + R_n(x). \tag{4}$$

We assume that $f(x)$ has a continuous derivative of order $(n + 1)$ in some region of variation of x including $x = a$, and we find $R_n(x)$

in terms of this derivative. Differentiating identity (4) once, twice, $\ldots n$ times, we have:

$$\left.\begin{aligned} f'(x) &= f'(a) + \frac{f''(a)}{1!}(x-a) + \ldots + \frac{f^{(n)}(a)}{(n-1)!}(x-a)^{n-1} + \\ &\qquad + R'_n(x)\,. \\ f''(x) &= f''(a) + \frac{f'''(a)}{1!}(x-a) + \ldots + \frac{f^{(n)}(a)}{(n-2)!}(x-a)^{n-2} + \\ &\qquad + R''_n(x)\,, \\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ f^{(n)}(x) &= f^{(n)}(a) + R_n^{(n)}(x)\,. \end{aligned}\right\} \qquad (4_1)$$

Putting $x = a$ in (4) and these last equations, we find:

$$R_n(a) = 0\,, \qquad R'_n(a) = 0\,, \qquad R_n^{(n)}(a) = 0\,. \qquad (5)$$

We find on again differentiating the last of equations (4_1):

$$R_n^{(n+1)}(x) = f^{(n+1)}(x)\,. \qquad (6)$$

We obtain an expression for $R_n(x)$ without difficulty from (5) and (6), since by a basic formula of the integral calculus:

$$R_n(x) - R_n(a) = \int_a^x R'_n(t)\,\mathrm{d}t\,,$$

whence, taking (5) into account and integrating by parts, we have:

$$R_n(x) = \int_a^x R'_n(t)\,\mathrm{d}t = -\int_a^x R'_n(t)\,\mathrm{d}(x-t) =$$

$$= -R'_n(t)(x-t)\Big|_a^x + \int_a^x R''_n(t)(x-t)\,\mathrm{d}t = -\int_a^x R''_n(t)\,\mathrm{d}\frac{(x-t)^2}{2!} =$$

$$= -R''_n(t)\frac{(x-t)^2}{2!}\Big|_a^x + \int_a^x R'''_n(t)\frac{(x-t)^2}{2!}\,\mathrm{d}t = -\int_a^x R'''_n(t)\,\mathrm{d}\frac{(x-t)^3}{3!} =$$

$$= -R'''_n(t)\frac{(x-t)^3}{3!}\Big|_a^x + \int_a^x R_n^{(4)}(t)\,\frac{(x-t)^3}{3!}\,\mathrm{d}t = \ldots =$$

$$= \int_a^x R_n^{(n+1)}(t)\,\frac{(x-t)^n}{n!}\,\mathrm{d}t = \frac{1}{n!}\int_a^x f^{(n+1)}(t)(x-t)^n\,\mathrm{d}t\,.$$

The above working is explained as follows. The variable of integration is denoted by t, so that x is to be taken as a constant under the integral sign, and the differential of x as zero; and hence, for example,

$$\mathrm{d}\,\frac{(x-t)^3}{3!} = \frac{3\,(x-t)^2}{3!}\,\mathrm{d}\,(x-t) = -\,\frac{(x-t)^2}{2!}\,\mathrm{d}t\,,$$

and in general,

$$\mathrm{d}\,\frac{(x-t)^k}{k!} = \frac{k\,(x-t)^{k-1}}{k!}\,\mathrm{d}\,(x-t) = -\,\frac{(x-t)^{k-1}}{(k-1)!}\,\mathrm{d}t.$$

The expression

$$R_n^{(k)}(t)\,\frac{(x-t)^k}{k!}\Big|_a^x \qquad (k \leqslant n)$$

is zero because the factor $(x-t)^k$ vanishes on substituting $t = x$ whilst the factor $R_n^{(k)}(a) = 0$ by (5) on substituting $t = a$.

We obtain in this way the following important proposition:

TAYLOR'S FORMULA. *Every function $f(x)$, having continuous derivatives up to and including order $(n+1)$ in some interval including $x = a$ as an interior point, can be expressed in powers of $(x-a)$ for all x inside the interval, as*

$$f(x) = f(a) + (x-a)\,\frac{f'(a)}{1!} + (x-a)^2\,\frac{f''(a)}{2!} + \ldots + \\ + (x-a)^n\,\frac{f^{(n)}(a)}{n!} + R_n(x)\,, \qquad (7)$$

where *the remainder term $R_n(x)$ has the form :*

$$R_n(x) = \frac{1}{n!}\int_a^x f^{(n+1)}(t)\,(x-t)^n\,\mathrm{d}t. \qquad (8)$$

A second form of the remainder term is often encountered in applications, this being derived at once from (8) by using the mean value theorem of [95]. The function $(x-t)^n$ under the integral on the right-hand side of (8) preserves its sign, so that we have by the mean value theorem:

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{n!}\int_a^x (x-t)^n\,\mathrm{d}t = \frac{f^{(n+1)}(\xi)}{n!}\left[-\,\frac{(x-t)^{n+1}}{n+1}\right]_a^x.$$

We obtain on substituting the upper and lower limits:

$$-\,\frac{(x-t)^{n+1}}{n+1}\Big|_a^x = \frac{(x-a)^{n+1}}{n+1}\,,$$

since the expression written vanishes for $t = x$. Substituting this in the above formula, we have:

$$R_n(x) = (x-a)^{n+1}\frac{f^{(n+1)}(\xi)}{(n+1)!}, \tag{9}$$

where ξ is a mean value lying between a and x. This result is known as *Lagrange's form of the remainder term,* and Taylor's formula with Lagrange's remainder reads:

$$f(x) = f(a) + (x-a)\frac{f'(a)}{1!} + (x-a)^2\frac{f''(a)}{2!} + \dots + \\ + (x-a)^n\frac{f^{(n)}(a)}{n!} + (x-a)^{n+1}\frac{f^{(n+1)}(\xi)}{(n+1)!} \tag{7_1}$$

(ξ between a and x).

127. Different forms of Taylor's formula. Putting $n = 0$ in (7_1) gives us Lagrange's formula for finite increments, deduced earlier in [63]:

$$f(x) - f(a) = (x-a)f'(\xi);$$

Taylor's formula is thus seen as a direct generalization of the formula for finite increments.

If we return to our previous notation and write x instead of a and $x + h$ instead of x, Taylor's formula (7) becomes:

$$f(x+h) - f(x) = \frac{hf'(x)}{1!} + \frac{h^2 f''(x)}{2!} + \dots + \frac{h^n f^{(n)}(x)}{n!} + R_n, \tag{10}$$

since $(x - a)$ is to be replaced by h in the new notation. The value ξ, lying between a and x in the previous notation, will lie between x and $(x + h)$, and can be written as $(x + \theta h)$, where $0 < \theta < 1$. Using (9), the remainder term of formula (10) can thus be written in the form:

$$R_n = h^{n+1}\frac{f^{(n+1)}(x+\theta h)}{(n+1)!} \qquad (0 < \theta < 1). \tag{11}$$

The left-hand side of (10) is the increment Δy of the function $y = f(x)$, corresponding to the increment, or what amounts to the same thing, to the differential h of the independent variable. Recalling the expressions for higher order differentials [55], we have:

$$dy = y'\,dx = f'(x) \times h, \ d^2 y = y''(dx)^2 = f''(x) \times h^2, \dots, \\ d^n y = y^{(n)}(dx)^n = f^{(n)}(x) \times h^n,$$

whence

$$\Delta y = \frac{\mathrm{d}y}{1!} + \frac{\mathrm{d}^2 y}{2!} + \ldots + \frac{\mathrm{d}^n y}{n!} + \left.\frac{\mathrm{d}^{n+1} y}{(n+1)!}\right|_{x+\theta h}, \tag{12}$$

the symbol:

$$\left.\frac{\mathrm{d}^{n+1} y}{(n+1)!}\right|_{x+\theta h}$$

denoting the result of substituting $x + \theta h$ instead of x in the expression:

$$\frac{\mathrm{d}^{n+1} y}{(n+1)!}.$$

Taylor's formula is of particular interest in this form when the increment h of the independent variable is an infinitesimal. The infinitesimal terms of different orders with respect to h in the increment Δy of the function can then be separated out by using (12).

The initial value a of the independent variable is commonly zero, in which case Taylor's formula (7) takes the form:

$$f(x) = f(0) + x\frac{f'(0)}{1!} + x^2\frac{f''(0)}{2!} + \ldots + x^n\frac{f^{(n)}(0)}{n!} + R_n(x), \tag{13}$$

where

$$R_n(x) = \frac{1}{n!}\int_0^x f^{(n+1)}(t)(x-t)^n\,\mathrm{d}t = \frac{x^{n+1} f^{(n+1)}(\xi)}{(n+1)!} = \frac{x^{n+1} f^{(n+1)}(\theta x)}{(n+1)!} \tag{14}$$

and ξ, lying between 0 and x, can be written: $\xi = \theta x$, where θ is a number satisfying $0 < \theta < 1$. Formula (13) is called *Maclaurin's formula.*

128. Taylor and Maclaurin series. If the given function $f(x)$ has derivatives of all orders, Taylor's and Maclaurin's formulae can be written for any n. We re-write (7) in the form:

$$f(x) - \left[f(a) + (x-a)\frac{f'(a)}{1!} + (x-a)^2\frac{f''(a)}{2!} + \ldots + (x-a)^n\frac{f^{(n)}(a)}{n!}\right] =$$

$$= f(x) - S_{n+1} = R_n(x),$$

where S_{n+1} is the sum of the first $(n+1)$ terms of the infinite series

$$f(a) + (x-a)\frac{f'(a)}{1!} + \ldots + (x-a)^n\frac{f^{(n)}(a)}{n!} +$$

$$+ (x-a)^{n+1}\frac{f^{(n+1)}(a)}{(n+1)!} + \ldots$$

If on indefinite increase of n:

$$\lim_{n\to\infty} R_n(x) = 0\,, \tag{15}$$

the series just written converges, by what was said in [118], and $f(x)$ is seen to be equal to the sum S of this series. We thus obtain *an expansion of the function $f(x)$ into an infinite Taylor power series*

$$f(x) = f(a) + (x-a)\frac{f'(a)}{1!} + \ldots + (x-a)^n \frac{f^{(n)}(a)}{n!} + \ldots \tag{16}$$

in powers of the difference $(x - a)$.

Maclaurin's formula similarly gives us, when condition (15) is satisfied:

$$f(x) = f(0) + x\frac{f'(0)}{1!} + \ldots + x^n \frac{f^{(n)}(0)}{n!} + \ldots \tag{17}$$

By evaluating the remainder term R_n as a function of n, we get the error involved in taking the sum of the first $(n + 1)$ terms of the series as $f(x)$ instead of the complete series. This is of great importance in the approximate evaluation of a function by means of its expansion into a power series, which is the method most commonly used in practice.

We apply the above working to the expansion and approximate evaluation of some simple functions.

129. Expansion of e^x. We have in the first place:

$$f(x) = e^x\,, \quad f'(x) = e^x\,, \;\ldots\,, \quad f^{(k)}(x) = e^x\,, \;\ldots\,,$$

so that

$$f(0) = f'(0) = \ldots = f^{(k)}(0) = 1\,,$$

and Maclaurin's formula with remainder term (14) gives:

$$f(x) = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \frac{x^{n+1}}{(n+1)!}e^{\theta x} \qquad (0 < \theta < 1)\,.$$

We have seen (Example [121]) that the series

$$\sum_{x=1}^{\infty} \frac{x^n}{n!}$$

is absolutely convergent for all finite x, and hence we have for all x:

$$\frac{x^{n+1}}{(n+1)!} \to 0 \quad \text{as} \quad n \to \infty\,,$$

since this expression is the general term of a convergent series.† On the other hand, the factor $e^{\theta x}$ in the expression for the remainder certainly does not exceed e^x for $x > 0$ and unity for $x < 0$, and the remainder therefore tends to zero for all x. This gives us the expansion:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + \ldots, \tag{18}$$

which is valid for all x.

In particular, putting $x = 1$, we get an expression for e which is very convenient for evaluating e to any degree of accuracy:

$$e = 1 + \frac{1}{1!} + \frac{1}{2!} + \ldots + \frac{1}{n!} + \ldots$$

We use this formula to evaluate e to six places of decimals. If we take the approximation:

$$e \sim 2 + \frac{1}{2!} + \ldots + \frac{1}{n!},$$

the error will be:

$$\frac{1}{(n+1)!} + \frac{1}{(n+2)!} + \ldots = \frac{1}{(n+1)!}\left[1 + \frac{1}{n+2} + \frac{1}{(n+2)(n+3)} + \ldots\right] <$$

$$< \frac{1}{(n+1)!}\left[1 + \frac{1}{n+1} + \frac{1}{(n+1)^2} + \ldots\right] = \frac{1}{(n+1)!} \times \frac{1}{1 - \dfrac{1}{n+1}} = \frac{1}{n!n},$$

where the ($<$) sign comes from replacing the factors $(n + 2)$, $(n + 3)$, $(n + 4)$, ... by the smaller numbers $(n + 1)$ in the denominator of the fraction, so that the fraction is increased.

We can thus say that e is included between the following limits:

$$2 + \frac{1}{2!} + \ldots + \frac{1}{n!} < e < 2 + \frac{1}{2!} + \ldots + \frac{1}{n!} + \frac{1}{n!n}.$$

If we want to find an approximate value for e, differing from the true value by not more than 0.000001, we put $n = 10$; then

$$e \sim 2 + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{10!},$$

and the error does not exceed $1/11!\,10 < 3 \times 10^{-8}$. The first two terms in this formula are calculated accurately; the remaining eight terms must be calculated to seven places, since the error of each term is then not greater than 0.5 of unity at the seventh place, i.e. 0.5×10^{-7}, and the total error is not greater than:

$$10^{-7} \times 0.5 \times 8 = 4 \times 10^{-7},$$

i.e. four times unity at the seventh place. Hence, the absolute value of the error as a whole does not exceed 4.3×10^{-7}.

† Cf. also the example in [34].

We have:

$2 = 2.000000$	0 (accurate)	
$\frac{1}{2!} = \frac{1}{2} = 0.500000$	0 ,,	
$\frac{1}{3!} = \frac{1}{2!\,3} = 0.166666$	7 (high value)	
$\frac{1}{4!} = \frac{1}{3!\,4} = 0.041666$	7 ,, ,,	
$\frac{1}{5!} = \frac{1}{4!\,5} = 0.008333$	3 (low value)	
$\frac{1}{6!} = \frac{1}{5!\,6} = 0.001388$	9 (high value)	$e \sim 2.7182818$
$\frac{1}{7!} = \frac{1}{6!\,7} = 0.000198$	4 (low value)	
$\frac{1}{8!} = \frac{1}{7!\,8} = 0.000024$	8 ,, ,,	
$\frac{1}{9!} = \frac{1}{8!\,9} = 0.000002$	8 (high value)	
$\frac{1}{10!} = \frac{1}{9!\,10} = 0.000000$	3 ,, ,,	

The value of e to 12 places is 2.718281828459.

130. Expansion of $\sin x$ and $\cos x$. We have [53]:

$$f(x) = \sin x, \quad f'(x) = \sin\left(x + \frac{1}{2}\pi\right), \quad \cdots, \quad f^{(k)}(x) = \sin\left(x + k\frac{1}{2}\pi\right),$$

whence

$$f(0) = 0, \quad f'(0) = 1, \quad f''(0) = 0, \quad f'''(0) = -1, \quad \cdots,$$
$$f^{(2m)}(0) = 0, \quad f^{(2m+1)}(0) = (-1)^m,$$

so that (13) now gives:

$$\sin x = \frac{x}{1!} - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + \frac{(-1)^n x^{2n+1}}{(2n+1)!} + \\ + \frac{x^{2n+3}}{(2n+3)!}\sin\left[\theta x + \frac{(2n+3)\pi}{2}\right].$$

The factor $x^{2n+3}/(2n+3)!$ tends to zero for $n \to \infty$, as we have seen above, whilst the absolute value of a sine does not exceed unity, so that the remainder tends to zero for all finite x, i.e. the expansion

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots + \frac{(-1)^n x^{2n+1}}{(2n+1)!} + \cdots \tag{19}$$

is valid *for all* x.

We can show similarly that the expansion

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \cdots + \frac{(-1)^n x^{2n}}{(2n)!} + \cdots \tag{20}$$

is valid *for all values of* x.

Series (19) and (20) are very convenient for calculating the values of $\sin x$ and $\cos x$ for small angles x. They are alternating for all positive or negative x, so that if we take such a number of terms that the later terms are decreasing, the absolute value of the error does not exceed the first of the neglected terms [123].

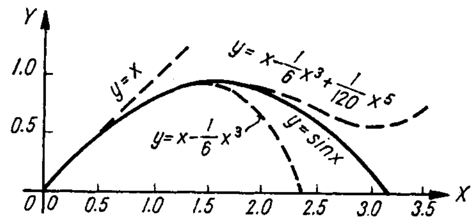

FIG. 156

Series (19) and (20) are only slowly convergent for large values of x, and are unsuitable for calculations. Figure 156 shows the relative disposition of the *accurate curve* of $\sin x$ and of the first three approximations:

$$x, \quad x - \frac{x^3}{6}, \quad x - \frac{x^3}{6} + \frac{x^5}{120}.$$

The more terms are taken in the approximation, the larger the interval in which the approximate is close to the accurate curve. We remark that angle x is expressed written in circular measure, i.e. in radians [33], in all the above formulae.

Example. To evaluate $\sin 10°$ to an accuracy of 10^{-5}. We first convert degrees to radians:

$$\text{arc } 10° = \frac{2\pi}{360} \times 10 = \frac{\pi}{18} = 0.17 \ldots$$

If we take the approximation

$$\sin \frac{\pi}{18} \sim \frac{\pi}{18} - \frac{1}{6}\left(\frac{\pi}{18}\right)^3,$$

our error does not exceed

$$\frac{1}{120} \times (0.2)^5 < 4 \times 10^{-6} \quad \left(\frac{\pi}{18} < 0.2\right).$$

Each term on the right-hand side of the above formula for $\sin \pi/18$ has to be evaluated to six places, since the total error will then not exceed

$$2 \times 0.5 \times 10^{-6} + 4 \times 10^{-6} = 5 \times 10^{-6}.$$

We have with the accuracy mentioned:

$$\frac{\pi}{18} = 0.174533\,; \quad \frac{1}{6}\left(\frac{\pi}{18}\right)^3 = 0.000886\,; \ \sin\frac{\pi}{18} = 0.173647\,,$$

where the first four places can be guaranteed.

131. Newton's binomial expansion. Here we have, taking $x > -1$, i.e. $1 + x > 0$:

$$f(x) = (1+x)^m,\ f'(x) = m(1+x)^{m-1},\ \ldots,$$

$$f^{(k)}(x) = m(m-1)\ldots(m-k+1)\,(1+x)^{m-k},$$

$$f(0) = 1,\ f'(0) = m,\ \ldots, f^{(k)}(0) = m(m-1)\ldots(m-k+1)\,,$$

where m is any real number; so that formula (13) gives us:

$$(1+x)^m = 1 + \frac{m}{1}x + \frac{m(m-1)\,x^2}{2!} + \ldots +$$

$$+ \frac{m(m-1)\,..\,(m-n+1)\,x^n}{n!} + R_n(x)\,, \tag{21}$$

where the remainder term can be found from (8) with $a = 0$:

$$R_n(x) = \frac{1}{n!}\int_0^x f^{(n+1)}(t)\,(x-t)^n\,\mathrm{d}t.$$

Noting that here:

$$f^{(n+1)}(t) = m(m-1)\ldots(m-n)\,(1+t)^{m-n-1},$$

we can write:

$$R_n(x) = \frac{m(m-1)\ldots(m-n)}{n!}\int_0^x (x-t)^n\,(1+t)^{m-n-1}\,\mathrm{d}t. \tag{22}$$

We apply to the integral the mean value theorem (13) of [95], letting θx, where $0 < \theta < 1$, denote the value of t, lying between 0 and x, that appears in this theorem; we get:

$$R_n(x) = \frac{m(m-1)\ldots(m-n)}{n!}(x-\theta x)^n(1+\theta x)^{m-n-1}\int_0^x \mathrm{d}t =$$

$$= \frac{(m-1)(m-2)\ldots(m-n)}{n!}\,x^n\left(\frac{1-\theta}{1+\theta x}\right)^n(1+\theta x)^{m-1}\,mx. \tag{23}$$

If $R_n \to 0$, the series

$$1 + \frac{mx}{1!} + \frac{m(m-1)}{2!}x^2 + \ldots + \frac{m(m-1)\ldots(m-n+1)}{n!}x^n + \ldots \tag{24}$$

must be convergent [118]. We have:

$$\left|\frac{u_{n+1}}{u_n}\right| = \left|\frac{m-n+1}{n}x\right| \to |x| \quad \text{for} \quad n \to \infty,$$

so that the series converges (absolutely) for $|x| < 1$ and diverges for $|x| > 1$ [124]. Although (24) converges for $|x| < 1$, it is still not clear that its sum is then $(1+x)^m$, and it has to be shown that $R_n(x) \to 0$ for $|x| < 1$. The factor

$$\frac{(m-1)(m-2)\ldots(m-n)}{n!}x^n$$

in expression (23) for $R_n(x)$ will be the general term of a *convergent* series (24) in which m has been replaced by $(m-1)$, and therefore [118] it tends to zero for $n \to \infty$.

The factor $[(1-\theta)/(1+\theta x)]^n$ does not exceed unity for any n. Also, we have here $-1 < x < +1$, so that $0 < 1 - \theta < 1 + \theta x$ for all positive or negative x, whence:

$$0 < \frac{1-\theta}{1+\theta x} < 1 \quad \text{and} \quad 0 < \left(\frac{1-\theta}{1+\theta x}\right)^n < 1.$$

The last factor $mx(1+\theta x)^{m-1}$ also remains bounded, since $(1+\theta x)$ lies between 1 and $1+x$, and $mx(1+\theta x)^{m-1}$ lies between mx and $mx(1+x)^{m-1}$, these limits being independent of n.

It is clear from these remarks that (23) gives $R_n(x)$ as the product of three factors, one of which tends to zero, whilst the other two are bounded on indefinite increase of n; and thus:

$$R_n(x) \to 0 \text{ for } n \to \infty.$$

Thus the expansion

$$(1+x)^m = 1 + \frac{mx}{1} + \frac{m(m-1)}{2!}x^2 + \ldots + \\ + \frac{m(m-1)\ldots(m-n+1)}{n!}x^n + \ldots \qquad (25)$$

is valid *for all values of x satisfying:*

$$|x| < 1.$$

When the exponent m is a positive integer, series (25) finishes at the term $n = m$ and reduces to Newton's elementary binomial formula. In the general case, the *expansion* (25) *is a generalization of Newton's binomial theorem for any exponent m.*

It is useful to note some special cases of the binomial expansion:

$$\frac{1}{1-x} = 1 + x + x^2 + x^3 + \ldots + x^n + \ldots, \tag{26}$$

$$\sqrt{1+x} = 1 + \frac{1}{2}x - \frac{1}{2\times 4}x^2 + \frac{1\times 3}{2\times 4\times 6}x^3 - \\ - \frac{1\times 3\times 5}{2\times 4\times 6\times 8}x^4 + \ldots, \tag{27}$$

$$\frac{1}{\sqrt{1+x}} = 1 - \frac{1}{2}x + \frac{1\times 3}{2\times 4}x^2 - \frac{1\times 3\times 5}{2\times 4\times 6}x^3 + \\ + \frac{1\times 3\times 5\times 7}{2\times 4\times 6\times 8}x^4 - \ldots \tag{28}$$

We remark that $(1+x)^m$ has a positive value for every $x > -1$ [19, 44], i.e. the sum of series (24) is positive for $-1 < x < +1$. In particular, for instance, series (27) gives the positive value of $\sqrt{1+x}$ in this interval.

Examples. **1.** *Extracting roots.* Formula (25) is specially suitable for extracting roots to any degree of accuracy. Let it be required to find the mth root of an integer A. An integer a can always be chosen so that its mth power is as close as possible to A; we put $A = a^m + b$, where $|b| < a^m$, which gives:

$$\sqrt[m]{A} = \sqrt[m]{a^m + b} = a\sqrt[m]{1 + \frac{b}{a^m}}.$$

Since here $|b/a^m| < 1$, then letting x denote b/a^m, we can calculate $\sqrt[m]{1+b/a^m}$ by Newton's binomial formula, the series being the better convergent, the smaller the absolute value of the ratio x.

We calculate, for instance, $\sqrt[5]{1000}$ to an accuracy of 10^{-5}. We have:

$$\sqrt[5]{1000} = \sqrt[5]{1024 - 24} = 4\left(1 - \frac{3}{128}\right)^{1/5} =$$

$$= 4\left[1 - \frac{1}{5}\times\frac{3}{128} - \frac{1}{5}\times\frac{4}{10}\times\left(\frac{3}{128}\right)^2 - \frac{1}{5}\times\frac{4}{10}\times\frac{9}{15}\left(\frac{3}{128}\right)^3 - \ldots\right].$$

We stop at the terms written and find the error, setting in (23):

$$m = \frac{1}{5}; \quad n = 3; \quad x = -\frac{3}{128}.$$

The factor $[(1-\theta)/(1+\theta x)]^n$ lies between zero and unity, as was shown. The factor $(1+\theta x)^{n-1}$ will be:

$$\left(1 - \theta\frac{3}{128}\right)^{-4/5} < \left(1 - \frac{3}{128}\right)^{-4/5} = \left(\frac{125}{128}\right)^{4/5} < \left(\frac{6}{5}\right)^{4/5} = \left(\sqrt[5]{\frac{6}{5}}\right)^4 < \left(\frac{4}{3}\right)^4,$$

since

$$\sqrt[3]{\frac{6}{5}} < \frac{6}{5} < \frac{4}{3}.$$

We get finally from (23):

$$4\,|R_n| < \frac{4}{1\times2\times3}\times\frac{1}{5}\times\frac{4}{5}\times\frac{9}{5}\times\frac{14}{5}\left(\frac{4}{128}\right)^4 <$$
$$< 2\times0.2\times0.8\times0.6\times2.8\times(0.03)^4 < 5\times10^{-7}.$$

Calculation of the other terms has to be carried out to six places, since the total error then will not be more than:

$$4\times3\times0.5\times10^{-6} + 5\times10^{-7} = 6.5\times10^{-6} < 10^{-5}.$$

The calculation can be set out as follows:

$\frac{1}{5}=0.2$	$\times\frac{3}{128} = 0.0234375 \times 0.2$	$=0.004687$
$\frac{1}{5}\times\frac{4}{10}=0.08$	$\times\left(\frac{3}{128}\right)^2 = 0.000549 \times 0.08$	$=0.000044$
$\frac{1}{5}\times\frac{4}{10}\times\frac{9}{15}=0.048$	$\times\left(\frac{3}{128}\right)^3 = 0.000013 \times 0.048$	$=0.000001$
		0.004732

$$\begin{array}{r} 1 - 0.004732 = 0.995268 \\ \times\, 4 \\ \hline 3.981072 \end{array}$$

2. *Approximate calculation of the length of an ellipse.* The following expression was obtained in [103] for the length l of an ellipse with semi-axes a and b:

$$l = 4\int_0^{\frac{1}{2}\pi}\sqrt{a^2\sin^2 t + b^2\cos^2 t}\,dt = 4a\int_0^{\frac{1}{2}\pi}\sqrt{\sin^2 t + \frac{b^2}{a^2}\cos^2 t}\,dt$$

[formula (22)]. Taking into account the eccentricity ε of the ellipse:

$$\varepsilon^2 = \frac{a^2-b^2}{a^2}, \quad \frac{b^2}{a^2} = 1-\varepsilon^2,$$

we get:

$$l = 4a\int_0^{\frac{1}{2}\pi}\sqrt{1-\varepsilon^2\cos^2 t}\,dt. \tag{29}$$

This integral cannot be found accurately, but can be evaluated to any desired degree of accuracy by expanding the integrand into a series of powers of ε:†

$$\sqrt{1-\varepsilon^2\cos^2 t} = 1 - \frac{1}{2}\varepsilon^2\cos^2 t + \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)}{1\times 2}\varepsilon^4\cos^4 t - $$

$$- \frac{\frac{1}{2}\left(\frac{1}{2}-1\right)\left(\frac{1}{2}-2\right)}{1\times 2\times 3}\varepsilon^6\cos^6 t + \ldots =$$

$$= 1 - \frac{1}{2}\varepsilon^2\cos^2 t - \frac{1}{8}\varepsilon^4\cos^4 t - \frac{1}{16}\varepsilon^6\cos^6 t + R_3,$$

where the error R_3, on evaluation by means of (23) with $n = 3$, satisfies the inequality:

$$|R_3| = \frac{\frac{1}{2}\times\frac{1}{2}\times\frac{3}{2}\times\frac{5}{2}}{1\times 2\times 3}\varepsilon^8\cos^8 t\left(\frac{1-\theta}{1-\theta\varepsilon^2\cos^2 t}\right)^3(1-\theta\varepsilon^2\cos^2 t)^{\frac{1}{2}-1} <$$

$$< \frac{5}{32}\frac{\varepsilon^8\cos^8 t}{\sqrt{1-\varepsilon^2}}, \tag{30}$$

since

$$0 < \left(\frac{1-\theta}{1-\theta\varepsilon^2\cos^2 t}\right)^3 < 1$$

and

$$(1-\theta^2\cos^2 t)^{\frac{1}{2}-1} < (1-\varepsilon^2\cos^2 t)^{-\frac{1}{2}}.$$

Substituting this expression in (29) for l, integrating, and recalling (27) [100], we find:

$$l = 4a\left[\int_0^{\frac{1}{2}\pi} dt - \frac{1}{2}\varepsilon^2\int_0^{\frac{1}{2}\pi}\cos^2 t\,dt - \frac{1}{8}\varepsilon^4\int_0^{\frac{1}{2}\pi}\cos^4 t\,dt - \frac{1}{16}\varepsilon^6\int_0^{\frac{1}{2}\pi}\cos^6 t\,dt + \int_0^{\frac{1}{2}\pi} R_3\,dt\right] =$$

$$= 2\pi a\left[1 - \frac{1}{4}\varepsilon^2 - \frac{3}{64}\varepsilon^4 - \frac{5}{256}\varepsilon^6 + \varrho\right], \tag{31}$$

where by (10_1) of [95] and inequality (30):

$$|\varrho| = \left|\frac{2}{\pi}\int_0^{\frac{1}{2}\pi} R^4\,dt\right| < \frac{5}{32}\frac{\varepsilon^8}{\sqrt{1-\varepsilon^2}}\frac{2}{\pi}\int_0^{\frac{1}{2}\pi}\cos^8 t\,dt = \frac{175}{2^{12}}\frac{\varepsilon^8}{\sqrt{1-\varepsilon^2}} < \frac{0.05\,\varepsilon^8}{\sqrt{1-\varepsilon^2}}.$$

† This expansion is in fact possible because $\varepsilon < 1$ for an ellipse, so that the term $-\varepsilon^2\cos^2 t$, having the role of x in Newton's binomial formula, is less than unity in absolute value.

Formula (31) is directly suitable for calculating the length of an ellipse, especially in the case of small eccentricity. It can also be used as the basis of a simple geometrical construction for finding approximately the length of an ellipse, using only ruler and compasses.

We let l_1 and l_2 respectively denote the arithmetic and geometric means of the semi-axes of the ellipse:

$$l_1 = \frac{a+b}{1}, \qquad l_2 = \sqrt{ab},$$

and we compare the length l of the ellipse with the circumferences $2\pi l_1$, $2\pi l_2$ of two circles of radii l_1 and l_2.

Noting that

$$b = a\sqrt{1-\varepsilon^2}, \qquad \frac{a+b}{2} = \frac{1}{2}a\left[1 - \sqrt{1-\varepsilon^2}\right], \qquad \sqrt{ab} = a\sqrt[4]{1-\varepsilon^2},$$

and expanding into series by Newton's binomial formula, we easily obtain the following expressions:

$$2\pi l_1 = 2\pi a\left[1 - \frac{1}{4}\varepsilon^2 - \frac{1}{16}\varepsilon^4 - \frac{1}{32}\varepsilon^6 + \varrho_1\right], \tag{32}$$

$$2\pi l_2 = 2\pi a\left[1 - \frac{1}{4}\varepsilon^2 - \frac{3}{32}\varepsilon^4 - \frac{7}{128}\varepsilon^6 + \varrho_2\right], \tag{33}$$

where the errors ϱ_1 and ϱ_2 satisfy the inequalities, on evaluating by (23)

$$|\varrho_1| < \frac{5}{32}\frac{\varepsilon^8}{\sqrt{1-\varepsilon^2}}, \qquad |\varrho_2| < \frac{77}{512}\frac{\varepsilon^8}{(1-\varepsilon^2)^{3/4}}$$

Hence it is clear that, with small eccentricity, *when higher powers of ε can be neglected compared with ε^2, the length of an ellipse can be taken as the circumference of a circle of radius equal to either the arithmetic or geometric mean of the semi-axes.* If greater accuracy is desired, we form the expression:

$$\alpha \cdot 2\pi l_1 + \beta \cdot 2\pi l_2, \tag{34}$$

α and β being chosen so that as many terms as possible of (31) and (34) coincide. Since the first two terms of each of (31), (32) and (33) coincide, we must have first of all

$$\alpha + \beta = 1.$$

Equating the coefficients of ε^4 in (31) and (34), we get further:

$$\frac{\alpha}{16} + \frac{3\beta}{32} = \frac{3}{64} \quad \text{or} \quad 4\alpha + 6\beta = 3.$$

Solving the equations obtained for α and β, we find:

$$\alpha = \frac{3}{2}, \quad \beta = -\frac{1}{2}.$$

Substituting this in (34), we have:

$$\alpha \cdot 2\pi l_1 + \beta \cdot 2\pi l_2 = 2\pi \left(\frac{3}{2} l_1 - \frac{1}{2} l_2\right) =$$

$$= 2\pi a \left(1 - \frac{1}{4}\varepsilon^2 - \frac{3}{64}\varepsilon^4 - \frac{5}{256}\varepsilon^6 + \frac{3}{2}\varrho_1 - \frac{1}{2}\varrho_2\right), \tag{35}$$

i.e. the terms in ε^6, as well as ε^4, are seen to coincide, and (31) and (35) only start to diverge with the term in ε^8. If we take into account the values found above for ϱ, ϱ_1 and ϱ_2, and if we note that

$$\frac{1}{\sqrt{1-\varepsilon^2}} \quad \text{and} \quad \frac{1}{(1-\varepsilon^2)^{3/4}} < \frac{1}{1-\varepsilon^2}, \quad \frac{175}{2^{12}} + \frac{5}{32} \times \frac{3}{2} + \frac{77}{512} \times \frac{1}{2} < 0.4,$$

we can finally say: *the error does not exceed* $0.4\varepsilon^8/(1-\varepsilon^2)$, *if the length of an ellipse with semi-axes a, b and eccentricity ε is taken as the circumference of a circle of radius r*, where:

$$r = \frac{3}{2}\,\frac{a+b}{2} - \frac{1}{2}\sqrt{ab}.$$

132. Expansion of $\log(1+x)$.† The general theory can be used for this expansion, but we shall employ a second method, which is suitable in many other cases.

We write $\log(1+x)$ in the form of a definite integral. We evidently have, for $x > -1$:

$$\int_0^x \frac{dt}{1+t} = \log(1+t)\Big|_0^x = \log(1+x) - \log 1 = \log(1+x),$$

i.e.

$$\log(1+x) = \int_0^x \frac{dt}{1+t}.$$

But we have the identity:

$$\frac{1}{1+t} = 1 - t + t^2 - t^3 + \ldots + (-1)^{n-1}t^{n-1} + \frac{(-1)^n t^n}{1+t},$$

† The function $\log x$ cannot be expanded into a series of powers of x, since the function itself and its derivatives have discontinuities at $x = 0$, where they tend to infinity.

which is immediately obtained on dividing unity by $1 + t$ and stopping at the remainder $(-1)^n t^n$. Thus:

$$\log(1+x) = \int_0^x \frac{dt}{1+t} =$$

$$= \int_0^x \left[1 - t + t^2 - t^3 + \ldots + (-1)^{n-1} t^{n-1} + \frac{(-1)^n t^n}{1+t}\right] dt =$$

$$= x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \ldots + \frac{(-1)^{n-1} x^n}{n} + R_n(x),$$

where

$$R_n(x) = (-1)^n \int_0^x \frac{t^n\, dt}{1+t}. \tag{36}$$

The series

$$x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots + \frac{(-1)^{n-1} x^n}{n} + \ldots,$$

for which

$$\left|\frac{u_n}{u_{n-1}}\right| = \frac{n-1}{n} |x| \to |x| \text{ for } n \to \infty,$$

is in fact divergent for $|x| > 1$ (Corollary [121]), and we can therefore only consider the cases:

$$|x| < 1, \; x = \pm 1.$$

Here, $x = -1$ has also to be discarded, since $\log(1 + x)$ becomes infinite for this value of x.

There remain the cases: (1) $|x| < 1$ and (2) $x = 1$. In case (1), applying the mean value theorem [95] to expression (36) for $R_n(x)$ and noting that t^n does not change sign as t varies from 0 to x, we obtain:

$$R_n(x) = \frac{(-1)^n}{1+\theta x} \int_0^x t^n\, dt = \frac{(-1)^n x^{n+1}}{(n+1)(1+\theta x)} \quad (0 < \theta < 1), \tag{37}$$

whence it follows that, since $|x| < 1$:

$$|R_n(x)| < \frac{1}{n+1} \times \frac{1}{1+\theta x}.$$

The factor $1/|1 + \theta x|$ on the right-hand side of the above inequality is bounded for all n, since it lies between the limits:

$$1 \text{ and } \frac{1}{1+x},$$

these being independent of n; so that for the values of x in question

$$R_n(x) \to 0 \quad \text{as} \quad n \to \infty .$$

We get the same result in case (2), when $x = 1$. The above formula (37) shows that, for $x = 1$:

$$|R_n(1)| = \frac{1}{n+1} \times \frac{1}{1+\theta} < \frac{1}{n+1},$$

i.e. again

$$R_n(1) \to 0 \quad \text{as} \quad n \to \infty .$$

The expansion:

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots + \frac{(-1)^{n-1}x^n}{n} + \dots \tag{38}$$

is thus valid *for all values of x, satisfying the inequalities:*

$$-1 < x \leqslant +1 . \tag{39}$$

In the particular case of $x = 1$, we have the equation:

$$\log 2 = 1 - \frac{1}{2} + \frac{1}{3} - \dots + \frac{(-1)^{n-1}}{n} + \dots ,$$

which has already been discussed [123]. Formula (38) is not suitable as it stands for computing logarithms, since it presupposes that x satisfies inequality (39); apart from which, the series on the right-hand side does not converge rapidly enough. It can be transformed into a more convenient form for computation. We replace x by $-x$ in the equation:

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \dots ,$$

which gives

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \dots \qquad (|x| < 1),$$

then we subtract term by term. We get:

$$\log\frac{1+x}{1-x} = 2\left(x + \frac{x^3}{3} + \frac{x^5}{5} \dots\right) \qquad (|x| < 1).$$

We put here:

$$\frac{1+x}{1-x} = 1 + \frac{z}{a} = \frac{a+z}{a}, \qquad x = \frac{z}{2a+z}, \tag{40}$$

and obtain:

$$\log\frac{a+z}{a} = 2\left[\frac{z}{2a+z} + \frac{1}{3} \times \frac{z^3}{(2a+z)^3} + \frac{1}{5} \times \frac{z^5}{(2a+z)^5} + \dots\right],$$

or

$$\log(a+z) = \log a + 2\left[\frac{z}{2a+z} + \frac{1}{3}\frac{z^3}{(2a+z)^3} + \dots\right] \qquad (41)$$

This formula is suitable for all positive values of a and z, since for these, $x = z/(2a + z)$ lies between zero and unity. The smaller the fraction $z/(2a + z)$, or what amounts to the same thing, the smaller z compared with a, the more suitable the formula is for computation.

Formula (41) is extremely useful for computing logarithms. The logarithmic tables were not in fact worked out by using series, which were still unknown at the time of Napier and Briggs; but still, (41) provides a handy means of checking or rapidly computing logarithmic tables. We put $z = 1$ in (41) and take the sequence:

$$a = 15,\ 24,\ 80,$$

which gives us:

$$\log 16 - \log 15 = 2\left[\frac{1}{31} + \frac{1}{3\times 31^3} + \dots\right] = 2P\,,$$

$$\log 25 - \log 24 = 2\left[\frac{1}{49} + \frac{1}{3\times 49^3} + \dots\right] = 2Q\,,$$

$$\log 81 - \log 80 = 2\left[\frac{1}{161} + \frac{1}{3\times 161^3} + \dots\right] = 2R\,,$$

where the series denoted by P, Q, R converge very rapidly. These equations lead us to the equations:

$$4\log 2 - \log 3 - \log 5 = 2P\,,$$

$$-3\log 2 - \log 3 + 2\log 5 = 2Q\,,$$

$$-4\log 2 + 4\log 3 - \log 5 = 2R\,,$$

which can easily be solved for:

$$\log 2,\ \log 3,\ \log 5\,,$$

giving us:

$$\log 2 = 14P + 10Q + 6R\,,$$

$$\log 3 = 22P + 16Q + 10R\,,$$

$$\log 5 = 32P + 24Q + 14R.$$

The logarithms obtained by this means will be natural; we use them to find the modulus M of the common system of logarithms:

$$M = \frac{1}{\log 10} = 0.4342945819\ \dots\,;$$

knowing this, we can pass from natural to common logarithms by means of the formula:

$$\log_{10} x = M \log x.$$

Similarly, factorizing as follows:

$$a = 2400 = 100\times 2^3\times 3, \qquad a + z = 2401 = 7^4,$$

$$a = 9800 = 100\times 2\times 7^2, \qquad a + z = 9801 = 3^4\times 11^2,$$

$$a = 123200 = 100\times 2^4\times 7\times 11, \qquad a + z = 123201 = 3^6\times 13^2,$$

$$a = 2600 = 100\times 2\times 13, \qquad a + z = 2601 = 3^2\times 17^2,$$

$$a = 28899 = 3^2\times 13^2\times 19, \qquad a + z = 28900 = 100\times 17^2,$$

we compute:

$$\log 7, \log 11, \log 13, \ldots$$

Having found the logarithms of the prime numbers, the evaluation of the logarithms of composite numbers follows simply by additions and multiplications by integers, without the aid of series, since, as we know, composite numbers can always be reduced to prime factors.

133. Expansion of arc tan x**.** We deal with this as in the case of $\log(1 + x)$. We have:

$$\mathrm{d}\,\mathrm{arc}\tan t = \frac{\mathrm{d}t}{1 + t^2}\cdot$$

We obtain by integrating:

$$\int_0^x \frac{\mathrm{d}t}{1 + t^2} = \mathrm{arc}\tan t\,\Big|_0^x = \mathrm{arc}\tan x - \mathrm{arc}\tan 0 = \mathrm{arc}\tan x\,,$$

where arc tan x takes its principal values, as in the example of [98]. We thus have:

$$\mathrm{arc}\tan x = \int_0^x \frac{\mathrm{d}t}{1+t^2} = \int_0^x \Big[1 - t^2 + t^4 - \ldots + (-1)^{n-1} t^{2n-2} +$$

$$+ \frac{(-1)^n t^{2n}}{1 + t^2}\Big]\mathrm{d}t = x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots + \frac{(-1)^{n-1} x^{2n-1}}{2n-1} + R_n(x)\,,$$

where

$$R_n(x) = (-1)^n \int_0^x \frac{t^{2n}\,\mathrm{d}t}{1 + t^2}\cdot \tag{42}$$

The series

$$x - \frac{x^3}{3} + \frac{x^5}{5} - \ldots + \frac{(-1)^{n-1} x^{2n-1}}{2n-1} + \ldots,$$

for which the ratio

$$\left|\frac{u_n}{u_{n-1}}\right| = \frac{2n-3}{2n-1}\,x^2 \to x^2 \quad \text{as} \quad n \to \infty\,,$$

is certainly divergent for $x^2 > 1$; so that we confine ourselves to the case of $x^2 \leqslant 1$, i.e.

$$-1 \leqslant x \leqslant +1. \tag{43}$$

Taking firstly $x > 0$, we get from (42), by VII [95]:

$$|R_n(x)| = \int_0^x \frac{t^{2n}}{1+t^2}\,dt < \int_0^x t^{2n}\,dt = \frac{x^{2n+1}}{2n+1} \leqslant \frac{1}{2n+1} \to 0 \qquad (n \to \infty),$$

since evidently

$$\frac{t^{2n}}{1+t^2} < t^{2n}.$$

If $x < 0$, we obtain on introducing a new variable in place of t, $t = -\tau$:

$$R_n(x) = (-1)^{n+1} \int_0^{-x} \frac{\tau^{2n}}{1+\tau^2}\,d\tau.$$

Here the upper limit $(-x)$ is positive, so that we have the same value as before for $R_n(x)$, i.e. the expansion

$$\operatorname{arc\,tan} x = x - \frac{x^3}{3} + \frac{x^5}{5} - \dots + \frac{(-1)^{n-1}x^{2n-1}}{2n-1} + \dots \tag{44}$$

is valid *for all x with absolute values not exceeding unity.*

In particular, we get for $x = 1$:

$$\operatorname{arc\,tan} 1 = \frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \dots$$

This series is unsuitable for computing π, in view of its very slow convergence. Series (44) converges the more rapidly, the smaller x. Let us put, for instance:

$$x = \frac{1}{5}, \quad \text{and} \quad \varphi = \operatorname{arc\,tan}\frac{1}{5}.$$

We have:

$$\tan 2\varphi = \frac{\frac{2}{5}}{1-\frac{1}{25}} = \frac{5}{12}, \quad \tan 4\varphi = \frac{\frac{5}{6}}{1-\frac{25}{144}} = \frac{120}{119}.$$

Since $\tan 4\varphi$ only differs slightly from unity, the angle 4φ is only slightly different from $\pi/4$. Let this small difference be:

$$\psi = 4\varphi - \frac{\pi}{4}, \quad \frac{\pi}{4} = 4\varphi - \psi.$$

Hence we deduce that:

$$\tan\psi = \tan\left(4\varphi - \frac{\pi}{4}\right) = \frac{\tan 4\varphi - \tan\frac{\pi}{4}}{1+\tan 4\varphi \times \tan\frac{\pi}{4}} = \frac{\frac{120}{119}-1}{1+\frac{120}{119}} = \frac{1}{239},$$

which gives:

$$\frac{\pi}{4} = 4\varphi - \psi = 4\operatorname{arc}\tan\frac{1}{5} - \operatorname{arc}\tan\frac{1}{239} =$$

$$= 4\left[\frac{1}{5} - \frac{1}{3}\times\frac{1}{5^3} + \frac{1}{5}\times\frac{1}{5^5} - \frac{1}{7}\times\frac{1}{5^7} + \dots\right] - \left[\frac{1}{239} - \dots\right].$$

Both series in brackets are alternating [123], so that if we confine ourselves to the terms written in each, our error will not exceed

$$\frac{4}{9\times 5^9} + \frac{1}{3\times 239^3} < 0.5\times 10^{-6}.$$

If π is required to an accuracy of 10^{-5}, we must compute the individual terms to seven places, since then the error for $\pi/4$ will not exceed:

$$4\times 4\times 0.5\times 10^{-7} + 0.5\times 10^{-7} + 0.5\times 10^{-6} < 2\times 10^{-6},$$

whilst the error for π will be less than 8×10^{-6}.
The computation can be set out as follows:

$\frac{1}{5} = 0.2000000$	$\frac{1}{3\times 5^3} = 0.0026667$	
$\frac{1}{5\times 5^5} = 0.0000640$	$\frac{1}{7\times 5^7} = 0.0000018$	
$+\,0.2000640$	$-\,0.0026685$	0.197 395 5
		$\times$ 4
		0.789 582 0
	$-\frac{1}{239} =$	$-$0.004 184 1
		0.785 397 9
		$\times$ 4
		$\pi \sim$ 3.141 591 6

The value of π to eight places is 3.14159165.
We can obtain the expansion, with $|x| \leqslant 1$:

$$\operatorname{arc}\sin x = \frac{x}{1} + \frac{1}{2}\frac{x^3}{3} + \frac{1\times 3}{2\times 4}\frac{x^5}{5} + \dots +$$

$$+ \frac{1\times 3\times 5 \dots (2n-1)}{2\times 4\times 6 \dots \quad 2n}\frac{x^{2n+1}}{2n+1} + \dots \tag{45}$$

134. Approximate formulae. Maclaurin's series, when convergent, enables a function $f(x)$ to be computed approximately by replacing

it by a finite number of terms of the expansion:

$$f(0) + \frac{xf'(0)}{1!} + \frac{x^2 f''(0)}{2!} + \dots$$

The smaller x, the smaller the number of terms that need be taken in this expansion in order to compute $f(x)$ with the desired accuracy. If x is very small, we can confine ourselves to the first two terms and neglect all the rest. We thus get *a very simple approximate formula for* $f(x)$, which, for small x, can entirely replace the often very complicated accurate expression for $f(x)$.

We note these approximate formulae for the most important functions:

$$\sqrt[n]{1+x} \sim 1 + \frac{x}{n}, \qquad \sin x \sim x,$$

$$\frac{1}{\sqrt[n]{1+x}} \sim 1 - \frac{x}{n}, \qquad \cos x \sim 1 - \frac{1}{2}x^2,$$

$$(1+x)^n \sim 1 + nx, \qquad \tan x \sim x,$$

$$a^x \sim 1 + x \log a, \qquad \log(1+x) \sim x.$$

By using these approximate formulae for (positive or negative) x near zero, complicated expressions can be very much simplified.

Examples.

1. $$\left(\frac{1+\frac{m}{n^2}x}{1-\frac{n-m}{n^2}x}\right)^n = \frac{\left(1+x\frac{m}{n^2}\right)^n}{\left(1-\frac{n-m}{n^2}x\right)^n} \sim \left(1+\frac{m}{n}x\right)\left(1-\frac{n-m}{n}x\right) \sim$$

$$\sim 1 + \frac{m}{n}x + \frac{n-m}{n}x = 1 + x.$$

2. $$\log\sqrt{\frac{1-x}{1+x}} = \frac{1}{2}\log(1-x) - \frac{1}{2}\log(1+x) \sim -\frac{1}{2}x - \frac{1}{2}x = -x.$$

3. To find the increase in volume of a substance on heating (cubical expansion), when the coefficient of linear expansion a is known. If one of the linear dimensions of the body is l_0 at $0°$, on heating to $t°$ it will be:

$$l = l_0(1 + at).$$

The coefficient of expansion a is very small for the majority of substances ($< 10^{-5}$). Since the ratio of the volumes is that of the cubes of the linear dimensions, we can write:

$$\frac{v}{v_0} = \frac{(1+at)^3}{1}; \quad v = v_0(1+at)^3 \sim v_0(1+3at),$$

i.e. the number $3a$ gives us the *coefficient of cubical expansion.* We find a similar relationship for the density ϱ, which is inversely proportional to the volume:

$$\frac{\varrho}{\varrho_0} = \frac{1}{(1 + at)^3}, \quad \varrho = \varrho_0 (1 + at)^{-3} \sim \varrho_0 (1 - 3\,at) .$$

All these approximate formulae are evidently only suitable for sufficiently small x; they prove quite inaccurate if this is not the case, and further terms of the expansion have to be taken into account.

135. Maxima, minima and points of inflexion. Taylor's formula allows an important addition to be made to the rules for finding the maxima and minima of a function, laid down in [58]. We shall assume in future that $f(x)$ has continuous derivatives up to order n at, and in the vicinity of, $x = x_0$.

If the first $(n - 1)$ *derivatives of* $f(x)$ *vanish at* $x = x_0$:

$$f'(x_0) = f''(x_0) = \ldots = f^{(n-1)}(x_0) = 0 ,$$

whilst the n-th derivative $f^{(n)}(x_0)$ *differs from zero, the value* x_0 *corresponds to a peak of the curve if* n, i.e. *the order of the first non-vanishing derivative, is even, in which case we have :*

$$\text{a maximum, if } f^{(n)}(x_0) < 0 ,$$

$$\text{a minimum, if } f^{(n)}(x_0) > 0 ;$$

whereas if n *is odd,* x_0 *corresponds to a point of inflexion, and not to a peak.*

We prove this by considering the differences:

$$f(x_0 + h) - f(x_0) \quad \text{and} \quad f(x_0 - h) - f(x_0) ,$$

where h is a suitably small positive number. By the definition of maximum and minimum [58], there will be a maximum at x_0 if both hese differences are less than zero, and a minimum if both are greater han zero. If these differences have the same sign for any arbitrarily mall positive h, there will be neither a maximum nor minimum at x_0. The differences are evaluated by substituting x_0 in place of a, and $\pm h$ in place of h in Taylor's formula:†

† We take Lagrange's form of the remainder term; the number θ, lying between zero and unity, is not the same for $(+h)$ as for $(-h)$, which is why we write θ_1 in the second formula.

$$f(x_0+h)=f(x_0)+\frac{h}{1!}f'(x_0)+\ldots+\frac{h^{n-1}}{(n-1)!}f^{(n-1)}(x_0)+$$
$$+\frac{h^n}{n!}f^{(n)}(x_0+\theta h),$$
$$f(x_0-h)=f(x_0)-\frac{h}{1!}f'(x_0)+\ldots+\frac{(-1)^{n-1}h^{n-1}}{(n-1)!}f^{(n-1)}(x_0)+$$
$$+\frac{(-1)^n h^n}{n!}f^{(n)}(x_0-\theta_1 h)\quad (0<\theta<1 \quad\text{and}\quad 0<\theta_1<1).$$

By hypothesis:

$$f'(x_0)=f''(x_0)=\ldots=f^{(n-1)}(x_0)=0,\quad f^{(n)}(x_0)\neq 0;$$

and therefore:

$$f(x_0+h)-f(x_0)=\frac{h^n}{n!}f^{(n)}(x_0+\theta h),$$
$$f(x_0+h)-f(x_0)=\frac{(-1)^n h^n}{n!}f^{(n)}(x_0-\theta_1 h).$$

For sufficiently small positive h, the factors:

$$f^{(n)}(x_0+\theta h)\ \text{and}\ f^{(n)}(x_0-\theta_1 h)$$

have, by the assumed continuity of $f^{(n)}(x)$, identical signs, the same, in fact, as that of the non-zero number $f^{(n)}(x_0)$.

We have seen that there can be a peak at x_0 when, and only when, both the differences:

$$f(x_0\pm h)-f(x_0)$$

have the same sign; but this can only be the case when n is even, by what has just been said; if n is odd, the factors h^n and $(-1)^n h^n$ have different signs, and the differences in question therefore also have different signs.

We now take n even; the differences

$$f(x_0\pm h)-f(x_0)$$

both have the sign of $f^{(n)}(x_0)$. If $f^{(n)}(x_0)<0$,

$$f(x_0\pm h)-f(x_0)<0,$$

and we have a maximum; whilst if $f^{(n)}(x_0)>0$,

$$f(x_0\pm h)-f(x_0)>0,$$

and we get a minimum.

If n is odd, Taylor's formula gives us for the second derivative $f''(x)$ for every case of $n \geqslant 3$:

$$f''(x_0 + h) = \frac{h^{n-2}}{(n-2)!} f^{(n)}(x_0 + \theta_2 h) ;$$

$$f''(x_0 - h) = \frac{(-1)^{n-2} h^{n-2}}{(n-2)!} f^{(n)}(x_0 - \theta_3 h),$$

whence we see, by a similar argument to the above, and noting that $(n - 2)$ is odd, that $f''(x)$ changes sign at its vanishing point $x = x_0$, i.e. x_0 corresponds to a point of inflexion [71], as it was required to prove.

136. Evaluation of indeterminate forms. We take the ratio of two functions

$$\frac{\varphi(x)}{\psi(x)}$$

which vanish at $x = a$. We evaluate the indeterminate form

$$\left.\frac{\varphi(x)}{\psi(x)}\right|_{x=a}$$

where $\varphi(a) = \psi(a) = 0$, by expanding numerator and denominator in accordance with Taylor's formula:

$$\varphi(x) = (x-a)\varphi'(a) + \frac{(x-a)^2}{2!}\varphi''(a) + \ldots + \frac{(x-a)^n \varphi^{(n)}(a)}{n!} + \\ + \frac{(x-a)^{n+1}\varphi^{(n+1)}(\xi_1)}{(n+1)!}$$

$$\psi(x) = (x-a)\psi'(a) + \frac{(x-a)^2}{2!}\psi''(a) + \ldots + \frac{(x-a)^n \psi^{(n)}(a)}{n!} + \\ + \frac{(x-a)^{n+1}\psi^{(n+1)}(\xi_1)}{(n+1)!}.$$

We then cancel out some power of $(x - a)$ in the ratio in question and afterwards put $x = a$.

Examples.

1. $$\lim_{x\to 0} \frac{1-\cos 2x}{e^{3x}-1-3x} = \lim_{x\to 0} \frac{1 - \left(1 - \frac{4x^2}{2} + \frac{16x^4}{24} + \ldots\right)}{\left(1 + 3x + \frac{9}{2}x^2 + \frac{27}{6}x^3 + \ldots\right) - 1 - 3x} =$$

$$= \lim_{x\to 0} \frac{2 - \frac{16x^2}{24} + \ldots}{\frac{9}{2} + \frac{27}{6}x + \ldots} = \frac{4}{9}.$$

The same method is useful in evaluating other types of indeterminate form. We take one example:

2. $\lim\limits_{x\to\infty}(\sqrt[3]{x^3-5x^2+1}-x)$.

Here we have the indeterminate form $(\infty - \infty)$. We have:

$$\sqrt[3]{x^3-5x^2+1}-x = x\left[\sqrt[3]{1-\frac{5x^2-1}{x^3}}-1\right] =$$
$$= x\left\{\left[1-\left(\frac{5}{x}-\frac{1}{x^3}\right)\right]^{1/3}-1\right\}.$$

For x sufficiently large in absolute value, the difference $\left(\frac{5}{x}-\frac{1}{x^3}\right)$ is near zero, and we can apply Newton's binomial formula with $m = 1/3$ and x replaced by $-(5/x - 1/x^3)$:

$$\left[1-\left(\frac{5}{x}-\frac{1}{x^3}\right)\right]^{1/3} = 1-\frac{1}{3}\left(\frac{5}{x}-\frac{1}{x^3}\right)+\frac{\frac{1}{3}\left(\frac{1}{3}-1\right)}{2!}\left(\frac{5}{x}-\frac{1}{x^3}\right)^2+\dots$$

Substituting this in the brackets above and cancelling the ones, we get:

$$\sqrt[3]{x^3-5x^2+1}-x = x\left[-\frac{1}{3}\left(\frac{5}{x}-\frac{1}{x^3}\right)+\frac{\frac{1}{3}\left(\frac{1}{3}-1\right)}{2!}\left(\frac{5}{x}-\frac{1}{x^3}\right)^2+\dots\right] =$$
$$= \left(-\frac{5}{3}+\frac{1}{3x^2}\right)+\dots,$$

where all the terms not written contain only negative powers of x, i.e. vanish in the limit as $x\to\infty$; hence:

$$\lim_{x\to\infty}\left(\sqrt[3]{x^3-5x^2+1}-x\right) = -\frac{5}{3}.$$

It is easy to justify the passages to the limit in infinite series, utilized in the present section, and we shall not dwell on this.

§ 14. Further remarks on the theory of series

137. Properties of absolutely convergent series. Absolutely convergent series were defined in [124]. We now establish their most important properties.

The sum of an absolutely convergent series is independent of the order of the terms.

We first prove this statement for series with non-negative terms, which, as we know [120], can only be either convergent (and hence absolutely convergent) or strictly divergent.

Let the convergent series

$$u_1 + u_2 + u_3 + \dots + u_n + \dots \tag{1}$$

be given, with positive (non-negative) terms.

Let s_n denote the sum of its first n terms, and s its sum. We evidently have:

$$s_n \leqslant s.$$

On rearranging the terms of (1) in any manner, we get a second distribution of terms, corresponding to the series

$$v_1 + v_2 + v_3 + \dots + v_n + \dots ; \tag{2}$$

each term of (1) has a definite place in (2), and conversely. Let σ_n denote the sum of the first n terms of series (2). For any n, a greater number m can be found, such that all the terms appearing in the sum σ_n come from s_m, and therefore

$$\sigma_n \leqslant s_m \leqslant s.$$

We have thus proved the existence of a constant number s, independent of n, such that, for all n:

$$\sigma_n \leqslant s.$$

it follows [120] that series (2) is convergent. Let its sum be σ. Evidently,

$$\sigma = \lim_{n\to\infty} \sigma_n \leqslant s.$$

Interchanging series (1) and (2) in the above argument, we can prove in the same way that

$$s \leqslant \sigma ,$$

and it follows from the inequalities $\sigma \leqslant s$, $s \leqslant \sigma$, that

$$s = \sigma.$$

We now turn to series with terms of any sign. Since series (1) is absolutely convergent by hypothesis, the series with positive terms

$$|u_1| + |u_2| + \dots + |u_n| + \dots = \sum_{n=1}^{\infty} |u_n| \tag{3}$$

Is convergent, and its sum s' is independent of the order of the terms by what has been proved. On the other hand, both the series

$$\sum_{n=1}^{\infty} \frac{1}{2}(|u_n| + u_n), \qquad \sum_{n=1}^{\infty} \frac{1}{2}(|u_n| - u_n)$$

(cf. [124]) also have positive terms and are also convergent, since the general term of each does not exceed $|u_n|$, i.e. the general term of the convergent series (3).

Their sums are independent of the order of the terms, by what has been proved; and the same can be said of their difference, which is identical with the sum of series (1). This was what we required to prove.

COROLLARY. *The terms of an absolutely convergent series can be grouped together in any manner and added in each group,* since the grouping is carried out by changing the order of the terms, which does not alter the sum of the series.

Note. If any sequence of terms is isolated from an absolutely convergent series, the series thus obtained is also absolutely convergent, since the operation corresponds to isolating a sequence of terms from series (3) with positive terms, which evidently does not affect the convergence of (3) and in fact reduces its sum. In particular, the series obtained by isolating the positive and negative terms of a convergent series are themselves convergent. Let s' denote the sum of the series consisting of positive terms, and $(-s'')$ the sum of the series consisting of negative terms. On indefinite increase of n, the sum s_n of the first n terms of the original series will contain as many terms as desired of the two series in question, and we obviously get in the limit:

$$s = \lim s_n = s' - s''.$$

It can easily be shown that, when a series is not absolutely convergent, the series made up of its positive terms, and that made up of its negative terms, are strictly divergent. For instance, the series [124]:

$$1 - \frac{1}{2} + \frac{1}{3} - \frac{1}{4} + \cdots$$

is not absolutely convergent; and the series

$$1 + \frac{1}{3} + \frac{1}{5} + \cdots \quad \text{and} \quad -\frac{1}{2} - \frac{1}{4} - \frac{1}{6} - \cdots$$

are divergent. The sum of the first n terms of the first series tends to $(+\infty)$, and that of the second series to $(-\infty)$, on indefinite increase of n. Riemann used the above-mentioned fact to show that, by suitably changing the order of the terms of a non-absolutely convergent series, its sum can be made equal to any desired number. It is thus seen that the concept of an absolutely convergent series is identical with the concept of a series, the sum of which does not depend on the order of the terms.

We note further that, if we change the places of a finite number of terms of a convergent (not necessarily absolutely convergent) series, the sum s_n of its first n terms remains the same for all sufficiently large n, i.e. the convergence of the series is unaffected, and its sum remains as before. The arguments and results above refer to the case when an infinite number of terms is rearranged.

138. Multiplication of absolutely convergent series. *The rule for multiplying finite sums can be used for cross-multiplying two absolutely convergent infinite series: their product is equal to the sum of the series which we ge tby multiplying each term of one series by each term of the other and adding the products obtained. The order of the terms is of no importance, since the series thus obtained is also absolutely convergent.*

Let the absolutely convergent series be given:

$$\left.\begin{aligned} s &= u_1 + u_2 + \ldots + u_n + \ldots \\ \sigma &= v_1 + v_2 + \ldots + v_n + \ldots \end{aligned}\right\} \tag{4}$$

We first take the special case of both series having positive terms, and also of multiplication being carried out in the following way:

$$u_1 v_1 + u_1 v_2 + u_2 v_1 + u_1 v_3 + u_2 v_2 + u_3 v_1 + \ldots + u_1 v_n + \\ + u_2 v_{n-1} + \ldots + u_n v_1 + \ldots \tag{5}$$

First of all, we show that series (5), all the terms of which are also positive, is convergent; then we show that its sum S is equal to $s\sigma$.

Let S_n denote the sum of the first n terms of (5). We can always select a large m, such that all the terms composing S_n come from the product of the sums:

$$s_m = u_1 + u_2 + \ldots + u_m, \quad \sigma_m = v_1 + v_2 + \ldots + v_m ,$$

i.e. so that $S_n \leqslant s_m \cdot s_n$, i.e.

$$S_n \leqslant s\sigma , \tag{6}$$

since $s_m \leqslant s$, $\sigma_m \leqslant \sigma$; hence follows the convergence of series (5) [120].

Letting S denote the sum of series (5), we evidently have, from inequality (6):

$$S = \lim_{n\to\infty} S_n \leqslant s\sigma .$$

We now consider the product $s_n \sigma_n$. Given n, a large m can evidently be found, such that all the terms comprising the product of the sums s_n and σ_n come from the sum S_m; we then get:

$$s_n\sigma_n \leqslant S_m \leqslant S ,$$

and therefore in the limit, as $n \to \infty$,

$$s_n\sigma_n \to s\sigma \leqslant S. \tag{7}$$

This inequality, together with (6), gives $S = s\sigma$, which it was required to prove.

Now let the series (4) be absolutely convergent, but with terms of arbitrary sign. We now have the convergence of the series with positive terms:

$$|u_1| + |u_2| + \ldots + |u_n| + \ldots \text{ and } |v_1| + |v_2| + \ldots + |v_n| + \ldots$$

and hence, by what has just been proved, the convergence of the series

$$|u_1||v_1| + |u_2||v_1| + |u_1||v_2| + |u_2||v_2| + \ldots + \\ + |u_1||v_n| + \ldots + |u_n||v_1| + \ldots$$

It is clear from this that series (5), composed in accordance with the above rule, is absolutely convergent in this case. We now let

$$a_1', \; a_2', \; \ldots, \; a_n', \; \ldots; \quad a_1'', \; a_2'', \; \ldots, \; a_n'', \; \ldots$$
$$b_1', \; b_2', \; \ldots, \; b_n', \; \ldots; \quad b_1'', \; b_2'', \; \ldots, \; b_n'', \; \ldots$$

denote respectively the positive terms of the series (4), and the absolute values of the negative terms. We know (cf. Note of [137]) that the series composed of these terms are convergent; we put

$$s' = \sum_{n=1}^{\infty} a_n', \; \sigma' = \sum_{n=1}^{\infty} b_n', \; s'' = \sum_{n=1}^{\infty} a_n'', \; \sigma'' = \sum_{n=1}^{\infty} b_n''. \tag{8}$$

We have [137]:

$$s = s' - s'', \; \sigma = \sigma' - \sigma''.$$

As has been shown, pairs of the series (8) with positive terms can be multiplied term by term; the sum of the products of the series

$$s'\sigma', \; s''\sigma'', \; -s'\sigma'', \; -s''\sigma'$$

contains those, and only those, terms which appear in series (5), and hence we have:

$$S = s'\sigma' + s''\sigma'' - s'\sigma'' - s''\sigma' = (s' - s'')(\sigma' - \sigma'')$$

which it was required to prove.

Example. The series

$$1 + q + q^2 + \ldots + q^{n-1} + \ldots = \frac{1}{1-q}$$

is absolutely convergent for $|q| < 1$, and therefore:

$$\frac{1}{(1-q)^2} = (1 + q + \ldots + q^{n-1} + \ldots)(1 + q + \ldots + q^{n-1} + \ldots) =$$
$$= 1 + 2q + 3q^2 + \ldots + nq^{n-1} + \ldots$$

139. Kummer's test. Cauchy's and d'Alembert's tests for the convergence and divergence of series [121], whilst of great practical importance, are extremely specialized, and cannot in fact be used in a number of fairly simple cases. The test given below possesses much greater generality.

Kummer's test. *The series with positive terms :*

$$u_1 + u_2 + \ldots + u_n + \ldots \tag{9}$$

is convergent, provided a sequence of positive numbers $a_1, a_2, \ldots, a_n, \ldots$ can be found, such that we always have, after some initial value of n :

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \geqslant a > 0, \tag{10}$$

where a is a positive number, independent of n ; series (9) *is divergent, provided*

we have, for the same values of n :

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} \leqslant 0, \tag{11}$$

where furthermore, the series $\sum\limits_{n=1}^{\infty} 1/a_n$ *is divergent.*

We can assume without loss of generality that the conditions of the theorem are satisfied, starting with $n = 1$. First, let condition (10) be satisfied. We deduce from this, putting $n = 1, 2, 3, \ldots$:

$$a_1u_1 - a_2u_2 \geqslant au_2,\ a_2u_2 - a_3u_3 \geqslant au_3,\ \ldots,\ a_{n-1}u_{n-1} - a_nu_n \geqslant au_n,$$

whence, adding term by term and cancelling as necessary, we find:

$$a(u_2 + \ldots + u_n) \leqslant a_1u_1 - a_nu_n < a_1u_1.$$

Thus, the sum of the first n terms of series (9) ecluding u_1 is less than a constant a_1u_1/a, independent of n; and therefore series (9) with positive terms is convergent [120].

Now let condition (11) be satisfied. It gives us:

$$\frac{u_{n+1}}{u_n} \geqslant \frac{\dfrac{1}{a_{n+1}}}{\dfrac{1}{a_n}},$$

i.e. u_{n+1}/u_n is not less than the corresponding ratio of the terms of the divergent series:

$$\sum_{n=1}^{\infty} \frac{1}{a_n}. \tag{12}$$

The divergence of series (9) now follows from the lemma below about series with positive terms:

Supplementary note on d'Alembert's test. *If, starting from a certain value of* n, u_{n+1}/u_n *does not exceed the corresponding ratio* v_{n+1}/v_n *of the terms of a convergent series* $\sum\limits_{n=1}^{\infty} v_n$, *the series* $\sum\limits_{n=1}^{\infty} u_n$ *is also convergent. Whereas, if* u_{n+1}/u_n *is not less than the corresponding ratio of the terms of a divergent series* $\sum\limits_{n=1}^{\infty} v_n$ *the series* $\sum\limits_{n=1}^{\infty} u_n$ *is also divergent.*

Suppose we have, in the first place:

$$\frac{u_{n+1}}{u_n} \leqslant \frac{v_{n+1}}{v_n},$$

where

$$\sum_{n=1}^{\infty} v_n \tag{13}$$

is convergent. We have consecutively:

$$\frac{u_n}{u_{n-1}} < \frac{v_n}{v_{n-1}}, \quad \frac{u_{n-1}}{u_{n-2}} < \frac{v_{n-1}}{v_{n-2}}, \ldots, \frac{u_2}{u_1} < \frac{v_2}{v_1}.$$

so that we find on cross-multiplying:

$$\frac{u_n}{u_1} \leqslant \frac{v_n}{v_1}, \quad \text{or} \quad u_n \leqslant \frac{u_1}{v_1} v_n.$$

The convergence of Σu_n follows from the last inequality and the remark in [120] (with $k = u_1/v_1$). The divergence of the series can be proved similarly in the case when $u_{n+1}/u_n \geqslant v_{n+1}/v_n$ and Σv_n is divergent.

140. Gauss's test. *Gauss's test* has extremely important applications.† *If in the series* (9) *with positive terms :*

$$u_1 + u_2 + \ldots + u_n + \ldots$$

the ratio u_n/u_{n+1} can be written in the form :

$$\frac{u_n}{u_{n+1}} = 1 + \frac{\mu}{n} + \frac{\omega_n}{n^p}, \quad \text{where} \quad p > 1 \quad \text{and} \quad |\omega_n| < A, \tag{14}$$

A being independent of n, i.e. *ω_n being bounded, series* (9) *is convergent for $\mu > 1$, and divergent for $\mu \leqslant 1$.*

We remark that d'Alembert's test is inapplicable in all the cases covered by this test [121]. Formula (14) is itself got by expanding u_n/u_{n+1} in powers of $1/n$, i.e. by sorting out the terms of different orders of smallness with respect to $1/n$, assuming, of course, that this is possible.

We turn to the proof, and investigate separately the cases: (1) $\mu \neq 1$, and (2) $\mu = 1$. In case (1), we put $a_n = n$ in Kummer's test, noting that $a_n > 0$ and that $\Sigma\, 1/n$ diverges [119]. We evidently have here:

$$\lim_{n\to\infty}\left[a_n \cdot \frac{u_n}{u_{n+1}} - a_{n+1}\right] = \lim_{n\to\infty}\left[n\left(1 + \frac{\mu}{n} + \frac{\omega_n}{n^p}\right) - n - 1\right] = \mu - 1.$$

If $\mu > 1$, we shall have, starting from a certain value of n:

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} > a > 0,$$

where a is any positive number less than $\mu - 1$; and series (9) will be convergent. Whereas if $\mu < 1$, we shall have, starting from a certain value of n:

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} < 0,$$

and series (9) will thus be divergent [139].

We have in case (2):

$$\frac{u_n}{u_{n+1}} = 1 + \frac{1}{n} + \frac{\omega_n}{n^p}.$$

† Actually, we are concerned with a generalization of the test established by Gauss.

We put $a_n = n \log n$ in Kummer's test, and form the series

$$\sum \frac{1}{a_n} = \sum \frac{1}{n \log n}, \tag{15}$$

where the summation can start at any positive integral n, since the first terms do not affect the convergence [118]. We prove the divergence of the series written, using Cauchy's integral test [122]. We must prove the divergence of the integral:

$$\int_a^\infty \frac{dx}{x \log x} \qquad (a > 1).$$

But we have:

$$\int_a^\infty \frac{dx}{x \log x} = \int_a^\infty \frac{d(\log x)}{\log x} = \int_{\log a}^\infty \frac{dt}{t} = \log(\log x)\Big|_a^\infty,$$

and $\log(\log x)$ increases indefinitely with increasing x, i.e. the integral written above is divergent; hence, series (15) is also divergent. We now form the difference $a_n(u_n/u_{n+1}) - a_{n+1}$, using (14):

$$a_n \frac{u_n}{u_{n+1}} - a_{n+1} = n\left(1 + \frac{1}{n} + \frac{\omega_n}{n^p}\right)\log n - (n+1)\log(n+1) =$$
$$= (n+1)\log n + \frac{\omega_n \log n}{n^{p-1}} - (n+1)\log(n+1) =$$
$$= \frac{\omega_n \log n}{n^{p-1}} + (n+1)\log\left(1 - \frac{1}{n+1}\right). \tag{16}$$

The factor ω_n is bounded by hypothesis, whilst $(\log n)/n^{p-1}$ tends to zero as $n \to \infty$, since $p - 1 > 0$ by hypothesis and $\log n$ increases more slowly than any positive power of n (example 2 of [66]). If we put $1/(n+1) = -x$, $x \to 0$, and the second term on the right becomes

$$(n+1)\log\left(1 - \frac{1}{n+1}\right) = -\frac{\log(1+x)}{x},$$

i.e. it tends to (-1) [38]. We thus see that $\Sigma\, 1/a_n$ diverges in this case, and $a_n(u_n/u_{n+1}) - a_{n+1} \to -1$ as $n \to \infty$, so that we have for sufficiently large n: $a_n(u_n/u_{n+1}) - a_{n+1} < 0$, i.e. series (9) is divergent [139], which is what we required to prove.

The convergence tests given above can be used for series with terms of any sign if u_n is replaced in them by $|u_n|$. But they only enable us to say in this case whether the given series is *absolutely convergent* or not. Generally speaking, we can elicit conditions of *absolute convergence* from them, but not conditions of *divergence*, because we know that series are not necessarily divergent when not absolutely convergent [124]. We thus obtain:

Supplementary note on Gauss's test. *The series*

$$u_1 + u_2 + \ldots + u_n + \ldots, \tag{17}$$

where the terms are positive or negative, and where

$$\left| \frac{u_n}{u_{n+1}} \right| = 1 + \frac{\mu}{n} + \frac{\omega_n}{n^p}, \tag{18}$$

with $p > 1$ and $|\omega_n| < A$, is absolutely convergent for $\mu > 1$.

The series is easily proved to be divergent for $\mu < 0$. We have in this case, taking into account the fact that ω_n is bounded:

$$\frac{\omega_n}{\mu n^{p-1}} \to 0, \quad 1 + \frac{\omega_n}{\mu n^{p-1}} \to 1 \quad \text{for} \quad n \to \infty,$$

and thus, starting from a certain value of n, since $\mu < 0$:

$$\frac{\mu}{n} + \frac{\omega_n}{n^p} = \frac{\mu}{n}\left(1 + \frac{\omega_n}{\mu n^{p-1}}\right) < 0 \quad \text{and} \quad \left| \frac{u_n}{u_{n+1}} \right| < 1,$$

i.e. starting from this n, the terms of the series increase in absolute values and the general term u_n cannot tend to zero as $n \to \infty$, i.e. series (17) is divergent.

141. Hypergeometric series. The above general considerations will be applied to the so-called *hypergeometric* or *Gaussian series* :

$$F(\alpha, \beta, \gamma\,;\, x) = 1 + \frac{\alpha\beta}{1!\gamma} x + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\gamma(\gamma+1)} x^2 + \ldots +$$
$$+ \frac{\alpha(\alpha+1)\ldots(\alpha+n-1)\beta(\beta+1)\ldots(\beta+n-1)}{n!\gamma(\gamma+1)\ldots(\gamma+n-1)} x^n + \ldots \tag{19}$$

Certain functions of applied mathematics lead to such series. The following particular cases can easily be verified by direct substitution for α, β, γ:

$$F(1, \beta, \beta\,;\, x) = 1 + x + x^2 + \ldots + x^n + \ldots = \frac{1}{1-x},$$
$$F(-m, \beta, \beta\,;\, x) = (1+x)^m,$$
$$\left.\frac{F(\alpha, \beta, \beta\,;\, -x) - 1}{\alpha}\right|_{\alpha=0} = \log(1+x). \tag{20}$$

We investigate the convergence of (19) by forming the ratio of two successive terms:

$$\frac{u_{n+1}}{u_n} = \frac{(\alpha+n)(\beta+n)}{(n+1)(\gamma+n)} x \to x \quad \text{as} \quad n \to \infty, \tag{21}$$

i.e. by the corollary of [121], (19) is convergent for $|x| < 1$, and divergent for $|x| > 1$. The only remaining cases are: (1) $x = 1$ and (2) $x = -1$. We also note that the factors $(\alpha + n)$, $(\beta + n)$, $(\gamma + n)$ are positive for all sufficiently large n, so that, for $x = 1$, all the terms of the series have the same sign for sufficiently large n, whilst we get an alternating series for large n when $x = -1$.

We expand by the progression formula in the first case (taking n sufficiently large) and cross-multiply term by term the resulting absolutely

convergent series [138]:

$$\frac{u_n}{u_{n+1}} = \frac{(n+1)(\gamma+n)}{(\alpha+n)(\beta+n)} = \frac{\left(1+\frac{1}{n}\right)\left(1+\frac{\gamma}{n}\right)}{\left(1+\frac{\alpha}{n}\right)\left(1+\frac{\beta}{n}\right)} =$$

$$= \left(1+\frac{1}{n}\right)\left(1+\frac{\gamma}{n}\right)\left(1-\frac{\alpha}{n}+\frac{\alpha^2}{n^2}-\frac{\alpha^3}{n^3}+\ldots\right)\left(1-\frac{\beta}{n}+\frac{\beta^2}{n^2}-\frac{\beta^3}{n^3}+\ldots\right) =$$

$$= 1 + \frac{\gamma-\alpha-\beta+1}{n} + \frac{\omega_n}{n^2},$$

where ω_n is bounded. Furthermore, if we neglect in this case a sufficiently large number of initial terms of the series

$$F(\alpha,\beta,\gamma;\ 1) = 1 + \frac{\alpha\beta}{1\cdot\gamma} + \ldots +$$
$$+ \frac{\alpha(\alpha+1)\ldots(\alpha+n-1)\beta(\beta+1)\ldots(\beta+n-1)}{n!\,\gamma(\gamma+1)\ldots(\gamma+n-1)} + \ldots,$$

we get a series with terms of constant sign; and on applying Gauss's test to this, we find *absolute convergence if*

$$\gamma-\alpha-\beta+1 > 1, \text{ i.e. } \gamma-\alpha-\beta > 0,$$

and divergence if

$$\gamma-\alpha-\beta+1 \leqslant 1, \quad \text{i.e.} \quad \gamma-\alpha-\beta \leqslant 0.$$

In the second case, when $x = -1$, we get an alternating series after a certain initial term:

$$1 - \frac{\alpha\beta}{1\cdot\gamma} + \frac{\alpha(\alpha+1)\beta(\beta+1)}{2!\,\gamma(\gamma+1)} - \ldots +$$
$$+ (-1)^n \frac{\alpha(\alpha+1)\ldots(\alpha+n-1)\beta(\beta+1)\ldots(\beta+n-1)}{n!\,\gamma(\gamma+1)\ldots(\gamma+n-1)} + \ldots$$

We have here, as earlier:

$$\left|\frac{u_n}{u_{n+1}}\right| = 1 + \frac{\gamma-\alpha-\beta+1}{n} + \frac{\omega_n}{n^2},$$

and hence, applying the supplementary note on Gauss's test, we find *convergence if*

$$\gamma-\alpha-\beta+1 > 1, \quad \text{i.e.} \quad \gamma-\alpha-\beta > 0,$$

and divergence if

$$\gamma-\alpha-\beta+1 < 0, \quad \text{i.e.} \quad \gamma-\alpha-\beta < -1.$$

In the case where

$$\gamma-\alpha-\beta = -1,$$

it can be shown that the general term of the series tends to a limit differing from zero, i.e. the series is *divergent* [119]. Finally, in the case when

$$-1 < \gamma-\alpha-\beta \leqslant 0,$$

it can be shown that the absolute value of the terms of the series is decreasing and tends to zero as $n \to \infty$, i.e. [123] the series is convergent, though not absolutely. We do not propose to dwell on the proof of these last two propositions.

If we apply this to the binomial expansion

$$(1+x)^m = 1 + \frac{m}{1!}x + \frac{m(m-1)}{2!}x^2 + \ldots + \\ + \frac{m(m-1)\ldots(m-n+1)}{n!}x^n + \ldots,$$

which is obtained from (19) ($\beta = \gamma$, arbitrary) by writing $(-m)$ in place of α and $(-x)$ in place of x, and which we know to be convergent for $|x| < 1$ and divergent for $|x| > 1$, we find that the series written is:

absolutely convergent	for $m > 0$,	if $x = -1$,
divergent	for $m < 0$,	if $x = -1$,
absolutely convergent	for $m > 0$,	if $x = 1$,
non-absolutely convergent	for $-1 < m < 0$,	if $x = -1$,
divergent	for $m > -1$,	if $x = 1$,

degenerating to a polynomial for $m =$ integer $\geqslant 0$.

We show later [149] that if a binomial series is convergent for $x = \pm 1$, its sum is $(1 \pm 1)^m$, i.e. 2^m or 0 respectively.

It must be noted that we have assumed above that α, β and γ are neither zero nor negative integers. This is important for γ, since otherwise the terms of the series become meaningless (the denominator vanishes), whilst if α or β is zero or a negative integer, the series breaks off and reduces to a finite sum.

142. Double series. We take a rectangular table of numbers, bounded above and to the left, but stretching to infinity on the right and downwards:

$$\begin{array}{c|cccccc}
 & 1 & 2 & 3 & \ldots & n & \ldots \\ \hline
1 & u_{11} & u_{12} & u_{13} & \ldots & u_{1n} & \ldots \\
2 & u_{21} & u_{22} & u_{23} & \ldots & u_{2n} & \ldots \\
3 & u_{31} & u_{32} & u_{33} & \ldots & u_{3n} & \ldots \\
 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
m & u_{m1} & u_{m2} & u_{m3} & \ldots & u_{mn} & \ldots \\
 & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots
\end{array} \tag{22}$$

The number of the *row* in the infinite set of rows is denoted by the first subscript of u, and similarly, the number of the *column* in the infinite set of columns is denoted by the second subscript. Thus, u_{ik} indicates the number at the intersection of the ith row and the kth column of the table.

Let us assume, in the first place, that all the u_{ik} are *positive*.

We define the sum of all the numbers of (22) by marking points with positive integral coordinates $M(i, k)$ in a coordinate plane and drawing a series of curves:

$$C_1, C_2, \ldots, C_n, \ldots,$$

intersecting the axes in the first quadrant, and subject only to the conditions that *every* point M falls inside the area (C_n), bounded by the curve C_n and the axes (Fig. 157), for sufficiently large n, and that area (C_n) lies inside (C_{n+1}). We form the sum S_n of all the numbers u_{ik}, corresponding to the points lying inside the area (C_n). This sum evidently increases as n increases,

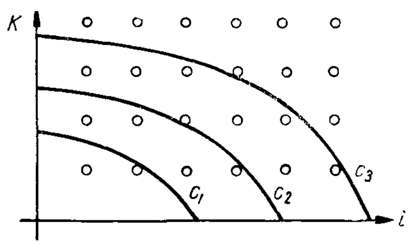

FIG. 157

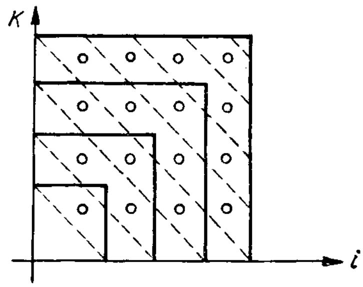

FIG. 158

so that only two cases are possible: either (1) S_n is bounded for all values of n, in which case there exists a finite limit:

$$\lim_{n\to\infty} S_n = S,$$

or (2) S_n increases indefinitely as n increases.

In case (1), *we say that the double series*

$$\sum_{i,k=1}^{\infty} u_{ik} \tag{23}$$

is convergent to the sum S. *In case* (2), *series* (23) *is said to be divergent.*

The sum of a convergent series (23) *with positive terms is independent of the method of summation*, i.e. *of the choice of curves* C_n, *and can be obtained by summation of the series in rows or in columns* :

$$S = \sum_{k=1}^{\infty} \left(\sum_{i=1}^{\infty} u_{ik} \right) = \sum_{i=1}^{\infty} \left(\sum_{k=1}^{\infty} u_{ik} \right), \tag{24}$$

i.e. we first find the sums of all the terms in each row (or each column of the table, then add the sums obtained.

Suppose, in fact, that we take any other system of curves $C_1, C_2, \ldots, C_n, \ldots$, with the same properties as $C'_1, C'_2, \ldots, C'_n, \ldots$ Let S'_n denote the sum of all the numbers of the table corresponding to points lying inside the area (C'_n). Given n, we can always choose a large m such that area (C'_n) appears

inside (C_m), in which case

$$S'_n < S_m < S\,,$$

i.e. there exists a finite limit:

$$\lim_{n\to\infty} S'_n = S' \leqslant S.$$

We can prove similarly, on interchanging the roles of the C'_n and C_n, that

$$S \leqslant S'\,,$$

which is only possible with

$$S = S'.$$

The sum of the double series (23) can be obtained even if we take as C_n step-lines formed from segments of straight lines (Fig. 158):

$$i = \text{const.},\ k = \text{const.}$$

This gives us a summation "by squares":

$$S = u_{11} + (u_{12} + u_{22} + u_{21}) + \dots + \\ + (u_{1n} + u_{2n} + \dots + u_{nn} + u_{n,n-1} + \dots + u_{n1}) + \dots$$

Summation "by diagonals" gives us:

$$S = u_{11} + (u_{12} + u_{21}) + (u_{13} + u_{22} + u_{31}) + \dots + \\ + (u_{1n} + u_{2,n-1} + \dots + u_{n1}) + \dots \tag{25}$$

We prove formula (24) by first of all noting that the sum of any number of terms of table (22) is less than S, and hence the sum of the terms in any given row, or in any given column, is also always less than S, whence follows the convergence of each of the series

$$\sum_{k=1}^{\infty} u_{ik} = s'_i\,, \qquad \sum_{i=1}^{\infty} u_{ik} = s''_k.$$

We have in addition, for any finite m and n:

$$\left.\begin{aligned} s'_1 + s'_2 + \dots + s'_m &= \sum_{i=1}^{m}\left(\sum_{k=1}^{\infty} u_{ik}\right) \leqslant S\,, \\ s''_1 + s''_2 + \dots + s''_n &= \sum_{k=1}^{n}\left(\sum_{i=1}^{\infty} u_{ik}\right) \leqslant S. \end{aligned}\right\} \tag{26}$$

We shall take, in fact, only the first m rows of table (22). We evidently have, on taking from these the elements of the first p columns:

$$\sum_{k=1}^{p}\left(\sum_{i=1}^{m} u_{ik}\right) < S.$$

We have by the rule for addition of series [119]:

$$s'_1 + s'_2 + \dots + s'_m = \sum_{k=1}^{\infty}\left(\sum_{i=1}^{m} u_{ik}\right) = \lim_{p\to\infty} \sum_{k=1}^{p}\left(\sum_{i=1}^{m} u_{ik}\right) \leqslant S\,,$$

since the expression under the limit sign is not greater than S.

The proof of the second of inequalities (26) is similar.

Inequalities (26) show that both the series:

$$\sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} u_{ik}\right) = \sum_{i=1}^{\infty} s_i' = \sigma', \quad \sum_{k=1}^{\infty}\left(\sum_{i=1}^{\infty} u_{ik}\right) = \sum_{k=1}^{\infty} s_k'' = \sigma''$$

are convergent, and have sums not exceeding S, i.e.

$$\sigma' \leqslant S \quad \text{and} \quad \sigma'' \leqslant S.$$

Obviously, on the other hand, for any choice of the system of curves C_r, all the terms that appear in the sum S_r appear in both the sums

$$s_1' + s_2' + \ldots + s_m', \quad s_1'' + s_2'' + \ldots + s_m''$$

with sufficiently large m, i.e.

$$S_r \leqslant s_1' + s_2' + \ldots + s_m' \leqslant \sigma', \quad S_r \leqslant s_1'' + \ldots + s_m'' \leqslant \sigma'',$$

and in the limit, therefore,

$$S = \lim_{r\to\infty} S_r \leqslant \sigma' \quad \text{and} \quad S \leqslant \sigma''.$$

Since $\sigma' \leqslant S$ and $\sigma'' \leqslant S$, we can only have

$$\sigma' = \sigma'' = S\,,$$

which it was required to prove.

As regards double series with terms of any sign (positive and negative), we only stop to consider *absolutely convergent series*, i.e. those *for which the double series consisting of absolute values:*

$$\sum_{i,k=1}^{\infty} |u_{ik}|$$

is convergent.

We can use an argument similar to that of [124] to show that *the sum :*

$$S = \lim_{n\to\infty} S_n = \sum_{i=1}^{\infty}\left(\sum_{k=1}^{\infty} u_{ik}\right) = \sum_{k=1}^{\infty}\left(\sum_{i=1}^{\infty} u_{ik}\right) \tag{27}$$

exists for such series, being independent of the method of summation, and being obtainable, in particular, by summation by rows and by columns.

Remark. Many properties of absolutely convergent simple series can be extended to absolutely convergent double series; in particular, the remark of [124] becomes: *if the terms of a double series do not exceed in absolute value the terms of a convergent double series with positive terms, the given series is absolutely convergent.*

Property (2) of [120] can be similarly extended.

Examples. **1.** The series

$$\sum_{i,k=1}^{\infty} \frac{1}{i^{\alpha}k^{\beta}} \tag{28}$$

is convergent for $\alpha > 1$, $\beta > 1$, since summation by squares gives:

$$S_n = \sum_{i=1}^{n}\left(\sum_{k=1}^{n}{}' \frac{1}{i^\alpha k^\beta}\right) = \left(\sum_{i=1}^{n}{}' \frac{1}{i^\alpha}\right)\left(\sum_{k=1}^{n}{}' \frac{1}{k^\beta}\right) < AB\,,$$

where A, B denote the sums of the series

$$\sum_{i=1}^{\infty}{}' \frac{1}{i^\alpha}\,, \quad \sum_{k=1}^{\infty}{}' \frac{1}{k^\beta}\,,$$

which are convergent for $\alpha > 1$, $\beta > 1$ [122].

2. The series

$$\sum_{i,k=1}^{\infty} \frac{1}{(i+k)^\alpha} \tag{29}$$

is convergent for $\alpha > 2$, and divergent for $\alpha \leqslant 2$, since summation by diagonals gives us:

$$S_n = \frac{1}{2^\alpha} + 2\cdot\frac{1}{3^\alpha} + \ldots + (n-1)\cdot\frac{1}{n^\alpha} =$$
$$= \frac{1}{2^{\alpha-1}}\left(1-\frac{1}{2}\right) + \ldots + \frac{1}{n^{\alpha-1}}\left(1-\frac{1}{n}\right),$$

whence we find, by replacing $(1 - 1/n)$, firstly by the smaller number $1/2$ secondly by the larger number 1:

$$\frac{1}{2}\left[\frac{1}{2^{\alpha-1}} + \ldots + \frac{1}{n^{\alpha-1}}\right] < S_n < \frac{1}{2^{\alpha-1}} + \ldots + \frac{1}{n^{\alpha-1}}\cdot$$

Our proposition now follows from the convergence of $\sum\limits_{n=1}^{\infty} 1/n^{\alpha-1}$ for $\alpha < 2$ and its divergence for $\alpha \leqslant 2$.

3. *If a and c are positive, and $b^2 - ac < 0$, the series*

$$\sum_{i,k=1}^{\infty} \frac{1}{(ai^2 + 2bik + ck^2)^p} \tag{30}$$

is convergent for $p > 1$ and divergent for $p \leqslant 1$.

We first take $b \geqslant 0$. Let A_1 denote the least of the numbers a and c, and A_2 the greatest of a, b, c; since obviously,

$$i^2 + k^2 \geqslant 2ik\,,$$

we have:

$$2A_1 ik \leqslant ai^2 + 2bik + ck^2 \leqslant A_2(i+k)^2\,,$$

whence we find, on confining our interest to the case $p > 0$:

$$\frac{1}{A_2^p}\,\frac{1}{(i+k)^{2p}} \leqslant \frac{1}{(ai^2 + 2bik + ck^2)^p} \leqslant \frac{1}{(2A_1)^p}\,\frac{1}{i^p k^p}\,;$$

noting that the factors $1/A_2^p$ and $1/(2A_1)^p$ are independent of i and k, this gives us convergence for $p > 1$ and divergence for $p \leqslant 1$, by Examples 1 and 2 and the remark made above.

Now take $b < 0$. Letting A_0 denote the greatest of a, c, $|b|$, we have by the obvious inequality $(\sqrt{ai})^2 + (\sqrt{ck})^2 > 2\sqrt{acik}$:

$$2(b + \sqrt{ac})\, ik \leqslant ai^2 + 2bik + ck^2 < A_0(i+k)^2 ,$$

where $b + \sqrt{ac} > 0$, since $|b| < \sqrt{ac}$ by hypothesis. The rest of the proof follows the same lines as for $b > 0$.

143. Series with variable terms. Uniformly convergent series. Taylor's and Maclaurin's formulae are examples of series, the terms of which depend on a variable x. In the second part of this course, we shall become familiar with the extremely important *trigonometric* series of the form:

$$\sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx) ,$$

the terms of which also depend on a variable x, as well as on n.

We now give a general discussion of series with terms that depend on an independent variable x.

Suppose we have an infinite sequence of functions:

$$u_1(x), u_2(x), \ldots, u_n(x), \ldots , \tag{31}$$

defined in an interval (a, b). *If the infinite series*

$$u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots \tag{32}$$

is convergent for all values of x in this interval, i.e. *for $a \leqslant x \leqslant b$, we say that it is convergent in the interval (a, b).*

The sum of the first n terms of series (32) is obviously a function of x, and similarly for the sum of the complete series and its remainder; we denote these respectively by

$$s_n(x),\ s(x),\ r_n(x) ,$$

so that

$$s(x) = \lim_{n\to\infty} s_n(x),\ r_n(x) = s(x) - s_n(x). \tag{33}$$

If (32) is convergent in (a, b) and has the sum $s(x)$, this means that, for any given x of (a, b), given any positive number ε, a number N can be found, such that

$$|r_n(x)| < \varepsilon \quad \text{for all} \quad n > N ,$$

with N obviously dependent on the choice of ε. But it must be noted that N will in general also depend on the x chosen, i.e. it can have different values for a single assigned ε and different values of x from (a, b); we indicate this by writing $N(x)$. *If for any given positive ε an N can be found, independent of x, such that for any x of (a, b) we have the inequality*

$$|r_n(x)| < \varepsilon \tag{34}$$

for all $n > N$, series (32) *is said to be uniformly convergent in the interval* (a, b).

We take as an example the series

$$\frac{1}{x+1} - \frac{1}{(x+1)(x+2)} - \frac{1}{(x+2)(x+3)} - \dots - \\ - \frac{1}{(x+n-1)(x+n)} - \dots, \tag{35}$$

where x varies in the interval $(0, a)$ and a is any given positive number.

It can easily be seen that the series can be written as:

$$\frac{1}{x+1} - \left(\frac{1}{x+1} - \frac{1}{x+2}\right) - \left(\frac{1}{x+2} - \frac{1}{x+3}\right) - \dots - \\ - \left(\frac{1}{x+n-1} - \frac{1}{x+n}\right) - \dots,$$

so that in this case

$$s_n(x) = \frac{1}{x+n}, \quad s(x) = \lim_{n\to\infty} s_n(x) = 0, \quad r_n(x) = -\frac{1}{x+n},$$

and if we want to make

$$|r_n(x)| = \frac{1}{x+n} < \varepsilon, \tag{36}$$

it is sufficient to take

$$n > \frac{1}{\varepsilon} - x = N(x). \tag{37}$$

If we now want to satisfy (36) for all x of $(0, a)$, given that $n > N$ where N is independent of the x taken, we need only put $N = 1/\varepsilon \geqslant N(x)$, since (37), and therefore (36), will then be satisfied for $n > N$ for all x of $(0, a)$. Series (35) is thus uniformly convergent in the interval $(0, a)$.

Not every series has the property of uniform convergence, since it is not possible to find an N independent of x for every series, such that it is not less than all $N(x)$ in interval (a, b).

Take, for instance, the series

$$x + x(x-1) + x^2(x-1) + \dots + x^{n-1}(x-1) + \dots \tag{38}$$

in the interval $0 \leqslant x \leqslant 1$.

The sum of its first n terms is:

$$s_n(x) = x + (x^2 - x) + (x^3 - x^2) + \dots + (x^n - x^{n-1}),$$

i.e.

$$s_n(x) = x^n,$$

and hence [26]:

$$s(x) = \lim_{n\to\infty} s_n(x) = 0 \qquad \text{for} \quad 0 \leqslant x < 1$$

and

$$r_n(x) = s(x) - s_n(x) = -x^n \quad \text{for} \quad 0 \leqslant x < 1.$$

For $x = 1$, substitution of $x = 1$ in (38) gives us the series

$$1 + 0 + 0 + \dots ,$$

i.e.

$$s_n(x) = 1 , \qquad s(x) = \lim_{n\to\infty} s_n(x) = 1 ,$$

$$r_n(x) = s(x) - s_n(x) = 0 ,$$

for $x = 1$ and any n. Series (38) is convergent throughout the interval $0 \leqslant x \leqslant 1$, but its convergence is not uniform in the interval. In fact, since $r_n(x) = -x^n$ for $0 \leqslant x < 1$, if we want to satisfy inequality (34), $| r_n(x) | < \varepsilon$, we must have

$$x^n < \varepsilon, \quad \text{i.e.} \quad n \log x < \log \varepsilon ,$$

or, dividing by the negative number $\log x$:

$$n > \frac{\log \varepsilon}{\log x} .$$

In this case, therefore, $N(x) = \log \varepsilon / \log x$ and cannot be replaced by a smaller number.

As x approaches unity, $\log x \to 0$, and the function $N(x)$ increases indefinitely; so that it is impossible to find an N, such that inequality (34) is satisfied for $n > N$ in the whole of (0,1). A consequence of this fact is that, whilst (38) is convergent throughout (0,1), including $x = 1$, its convergence becomes slower as x approaches unity; the nearer x is to unity, the more terms of the series have to be taken in order to approach its sum. Yet it is to be noted that the series simply breaks off at the second term when, in fact, $x = 1$.

We now give a second definition of uniform convergence, equivalent to the first. We formulated above [125] a necessary and sufficient condition for the convergence of a series. This now becomes: a necessary and sufficient condition for the convergence of series (32) in the interval (a, b) is that, for any given positive ε and for any x of (a, b), there exists an N, such that

$$| u_{n+1}(x) + u_{n+2}(x) + \dots + u_{n+p}(x) | < \varepsilon \tag{39}$$

for $n > N$ and any positive integer p. For a given ε, this N can still depend on the choice of x. *If for any given positive ε there exists the same N for all x of (a, b), such that* (39) *is satisfied for $n > N$ and any positive integer p, series* (32) *is said to be uniformly convergent in the interval* (a, b).

We have to show that this new definition of uniform convergence is equivalent to the first, i.e. if a series is uniformly convergent in the first sense, it is uniformly convergent in the new sense, and vice versa. To start with, let the series be uniformly convergent in the first sense, i.e. $| r_n(x) | < \varepsilon$ for $n > N$, where x is any x of (a, b) and N does not depend on x. We evidently have:

$$u_{n+1}(x) + u_{n+2}(x) + \dots + u_{n+p}(x) = r_n(x) - r_{n+p}(x) \tag{40}$$

and therefore:

$$| u_{n+1}(x) + u_{n+2}(x) + \dots + u_{n+p}(x) | \leqslant | r_n(x) | + | r_{n+p}(x) | .$$

which gives, for $n > N$, and so $n + p > N$:

$$|u_{n+1}(x) + u_{n+2}(x) + \ldots + u_{n+p}(x)| < 2\varepsilon. \tag{41}$$

In view of the arbitrary choice of ε, we see that the series is uniformly convergent in the new sense. We now start from the assumption that the series is uniformly convergent in the new sense, i.e. that (39) is satisfied for $n > N$, where N is independent of x, for any positive integer p and for any x of (a, b). Noting that

$$r_n(x) = u_{n+1}(x) + u_{n+2}(x) + \ldots = \lim_{p \to \infty} [u_{n+1}(x) + \ldots + u_{n+p}(x)]$$

we obtain in the limit, as $p \to \infty$ in (39):

$$|r_n(x)| \leqslant \varepsilon$$

for $n > N$, i.e. since ε is arbitrary, the first definition of uniform convergence follows from the new definition. The equivalence of the two definitions is thus proved.

We remark that $r_n(x)$ was used in the first definition (34) of uniform convergence and, at the same time, the additional assumption was made that the series was convergent; whereas, the fact of the convergence of the series is contained in the new definition of uniform convergence, of (39).

144. Uniformly convergent sequences of functions. The sequence of functions:

$$s_1(x), s_2(x), \ldots, s_n(x), \ldots, \tag{42}$$

which we considered above, was defined with the aid of series (32); $s_n(x)$ denoted the sum of the first n terms of the series. But sequence (42) can be taken as given and can be considered on its own account, and a series constructed from it, the sum of the first n terms of which is the nth term of the sequence $s_n(x)$. The terms of this series are obviously defined by:

$$u_1(x) = s_1(x), u_2(x) = s_2(x) - s_1(x), \ldots, u_n(x) = s_n(x) - s_{n-1}(x), \ldots \tag{43}$$

Sequence (42) is often simpler than (43), as was the case in the examples considered.

We arrive in this way at the concepts of convergence and uniform convergence of sequences of functions:

If a sequence of functions (42):

$$s_1(x), s_2(x), \ldots, s_n(x), \ldots,$$

is defined in an interval (a, b), *and if for any* x *of this interval the limit exists:*

$$s(x) = \lim_{n \to \infty} s_n(x), \tag{44}$$

sequence (42) *is said to be convergent in the interval* (a, b), *whilst* $s(x)$ *is called the limit function of sequence* (42).

If, in addition, for any previously assigned positive ε *there exists a number* N, *independent of* x, *such that the inequality*

$$|s(x) - s_n(x)| < \varepsilon \tag{45}$$

is valid for every $n > N$ throughout (a, b), sequence (42) is said to be uniformly convergent in the interval (a, b). Condition (45) may be replaced by the equivalent

$$| s_m(x) - s_n(x) | < \varepsilon \tag{46}$$

for m and $n > N$.

The condition for uniform convergence of sequence (42) is equivalent to the condition for uniform convergence of the series

$$u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots , \tag{47}$$

where (43)

$$u_1(x) = s_1(x), u_2(x) = s_2(x) - s_1(x), \ldots, u_n(x) = s_n(x) - s_{n-1}(x), \ldots$$

The proof of the equivalence of conditions (45) and (46) in the study of the uniform convergence of sequences is similar to that of the equivalence of conditions (35) and (36) for infinite series. We also note that the uniform convergence of $s_n(x)$ in any part of (a, b) follows at once from its uniform convergence in (a, b).

The uniform convergence of sequences can be interpreted geometrically. If $s(x)$ and $s_n(x)$ are illustrated graphically for different values of n, uniform convergence of the sequence implies that the piece of ordinate, comprised between the curves $s_n(x)$ and $s(x)$, must tend to zero as $n \to \infty$, for all x of (a, b); this condition is not fulfilled with non-uniform convergence of the sequence.

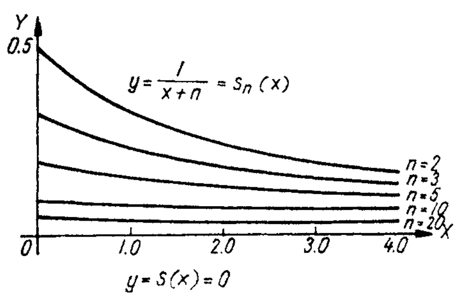

FIG. 159

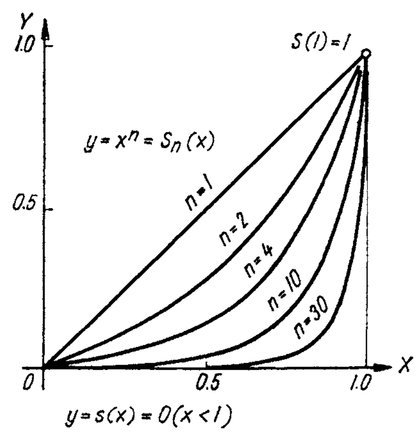

FIG. 160

The situation can be seen by inspection of Figs. 159 and 160, drawn for the examples that were discussed above:

$$s_n(x) = \frac{1}{x+n}, \quad s_n(x) = x^n. \dagger$$

† Different scales are used in Figs. 159 and 160 for x and y, for greater clarity.

In the case of Fig. 160, the limit function $s(x)$ is represented graphically by the segment (0,1) of axis OX, excluding the point 1, and by the isolated point with coordinates (1,1).

Admittedly, the limit function $s(x)$ in the last example is not continuous; but an example can easily be found of a sequence which, whilst having a continuous limit function, is nevertheless non-uniformly convergent.

The sequence (Fig. 161):

$$s_n(x) = \frac{nx}{1+n^2x^2} \quad (0 \leqslant x \leqslant a) \quad (48)$$

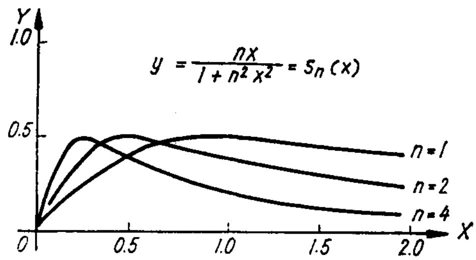

FIG. 161

has this property. We evidently have, for $x \neq 0$:

$$\frac{nx}{1+n^2x^2} = \frac{1}{n} \cdot \frac{x}{\frac{1}{n^2}+x^2},$$

and as $n \to \infty$, the first factor on the right, $1/n \to 0$, whilst the second tends to $1/x$, i.e. $s_n(x) \to 0$ for $x \neq 0$. Obviously, for $x = 0$, $s_n(0) = 0$ for all n and we therefore have, for all x of $(0, a)$, where a is a given positive number:

$$s(x) = \lim_{n\to\infty} s_n(x) = 0.$$

But the maximum piece of ordinate between curves $s_n(x)$ and $s(x)$, which simply reduces in the present case to the ordinate of $s_n(x)$, since $s(x) = 0$, is $1/2$ (corresponding to $x = 1/n$). Since it does not tend to zero as $n \to \infty$, sequence (48) is not uniformly convergent in $(0, a)$. In fact, if we want to have:

$$|s(x) - s_n(x)| = \frac{nx}{1+n^2x^2} < \varepsilon,$$

solving the second degree inequality:

$$0 < 1 - \frac{x}{\varepsilon} n + x^2n^2$$

for n, and taking ε sufficiently small, we get:

$$n > \frac{1}{2x\varepsilon}\left(\sqrt{1-4\varepsilon^2}\right) = N(x).$$

This function increases indefinitely as $x \to 0$, which accounts for the non-uniform convergence of the sequence.

We remark finally that Figs. 160 and 161 also show that the sequence x^n is uniformly convergent in an interval $(0, q)$, where q is any positive number less than unity, whilst the sequence $nx/(1+n^2x^2)$ is uniformly convergent in (q, a), where $0 < q < a$; a direct proof also of these statements may also easily be given.

145. Properties of uniformly convergent sequences. 1. *The limit function of a sequence of continuous functions, uniformly convergent in an interval* (a, b), *is also continuous.* Let

$$s_1(x), s_2(x), \ldots, s_n(x), \ldots$$

be the given sequence of functions, all of which are continuous in (a, b), and let

$$s(x) = \lim_{n\to\infty} s_n(x)$$

be the limit function. We have to show that, for any previously assigned small positive ε, a δ can be found, such that [35]:

$$|s(x + h) - s(x)| < \varepsilon \tag{49}$$

if

$$|h| < \delta,$$

assuming that both x and $x + h$ lie in (a, b). We can write for any n:

$$|s(x + h) - s(x)| =$$

$$= |[s(x + h) - s_n(x + h)] + [s_n(x + h) - s_n(x)] + [s_n(x) - s(x)]| \leqslant$$

$$\leqslant |s(x + h) - s_n(x + h)| + |s(x) - s_n(x)| + |s_n(x + h) - s_n(x)|.$$

By the definition of uniform convergence, we can select a suitably large n so that, throughout (a, b), including the points x and $x + h$:

$$|s(x + h) - s_n(x + h)| < \frac{\varepsilon}{3}, \quad |s(x) - s_n(x)| < \frac{\varepsilon}{3}.$$

Having fixed such an n, by the continuity of $s_n(x)$ [35], we can find a δ such that

$$|s_n(x + h) - s_n(x)| < \frac{\varepsilon}{3}, \quad \text{if} \quad |h| < \delta.$$

We combine all these inequalities, to obtain inequality (49).

If the sequence of functions is non-uniformly convergent, the possibility arises of the limit function not being continuous; the sequence x^n in the interval (0,1) can be quoted as an instance of this.

The converse does not hold, however: the limit function of a non-uniformly convergent sequence may be continuous, as, for instance, in the case of the sequence:

$$\frac{nx}{1 + n^2x^2}.$$

2. *If*

$$s_1(x), s_2(x), \ldots, s_n(x), \ldots$$

is a uniformly convergent sequence of functions, continuous in the interval (a, b), *and* (α, β) *is any interval lying in* (a, b),

$$\int_\alpha^\beta s_n(x)\,dx \to \int_\alpha^\beta s(x)\,dx \tag{50}$$

or alternatively,

$$\lim_{n\to\infty} \int_a^\beta s_n(x)\,dx = \int_a^\beta \lim_{n\to\infty} s_n(x)\,dx. \tag{51}$$

If the limits of integration are variable, e.g., $\beta = x$, *the sequence of functions*

$$\int_a^x s_n(t)\,dt \qquad (n = 1, 2, 3, \ldots) \tag{52}$$

is also uniformly convergent in the interval (a, b). (This process is referred to as *passing to the limit under the integral sign.*)

We note, first of all, that the limit function $s(x)$ is also continuous, by property (1). We now take the difference:

$$\int_a^\beta s(x)\,dx - \int_a^\beta s_n(x)\,dx = \int_a^\beta [s(x) - s_n(x)]\,dx.$$

Given ε, in view of the uniform convergence, we can find an N such that, for all $n > N$, we have throughout (a, b):

$$|\, s(x) - s_n(x) \,| < \varepsilon,$$

and hence [95] (10_1):

$$\left| \int_a^\beta [s(x) - s_n(x)]\,dx \right| < \int_a^\beta |\, s(x) - s_n(x) \,|\,dx < \int_a^\beta \varepsilon dx = \varepsilon(\beta - a) \leqslant \varepsilon(b - a).$$

We therefore have, for any interval (a, β) contained in (a, b):

$$\left| \int_a^\beta s(x)\,dx - \int_a^\beta s_n(x)\,dx \right| < \varepsilon(b - a) \quad \text{for} \quad n > N.$$

The right-hand side of this inequality is independent of a and β and tends to zero if $\varepsilon \to 0$. In view of the arbitrariness of ε, we can state the result thus: for any given positive ε_1 there exists an N, independent of a and β, such that

$$\left| \int_a^\beta s(x)\,dx - \int_a^\beta s_n(x)\,dx \right| < \varepsilon_1$$

for $n > N$. Formula (50) immediately follows from this. Putting $\beta = x$, and remembering that N is independent of β, we see that sequence (52) is uniformly convergent for all x of (a, b).

The theorem can be proved false for a non-uniformly convergent sequence. For instance, let:

$$s_n(x) = nxe^{-nx^2} \qquad (0 \leqslant x \leqslant 1)$$

(Fig. 162). It can easily be shown, taking the cases $x > 0$ and $x = 0$ separately, that

$$s_n(x) \to 0 \quad \text{as} \quad n \to \infty$$

for all x of (0,1), so that here, $s(x) = 0$. The sequence cannot be uniformly convergent, however, since the greatest ordinate of the curve $y = s_n(x)$ or, what is the same thing, the greatest difference $s_n(x) - s(x)$, which is obtained with $x = 1/\sqrt{2n}$, increases indefinitely as $n \to \infty$.

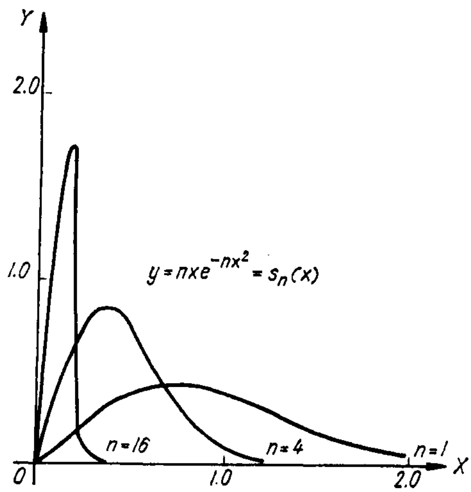

FIG. 162

On the other hand, we have:

$$\int_0^1 s_n(x)\,dx = n\int_0^1 xe^{-nx^2}\,dx =$$

$$= -\frac{1}{2}e^{-nx^2}\Big|_0^1 = \frac{1}{2}(1 - e^{-n}) \to \frac{1}{2},$$

and at the same time:

$$\int_0^1 s(x)\,dx = 0.$$

3. *If the functions of a sequence :*

$$s_1(x), s_2(x), \ldots, s_n(x), \ldots$$

have continuous derivatives :

$$s_1'(x), s_2'(x), \ldots, s_n'(x), \ldots$$

in an interval (a, b), and if the sequence $s_n'(x)$ is uniformly convergent to the limit function $\sigma(x)$, and sequence $s_n(x)$ convergent to $s(x)$, $s_n(x)$ is then also uniformly convergent and:

$$\sigma(x) = \frac{ds(x)}{dx}, \tag{53}$$

or alternatively :

$$\lim_{n\to\infty} \frac{ds_n(x)}{dx} = \frac{d \lim\limits_{n\to\infty} s_n(x)}{dx}. \tag{54}$$

This process is called *passing to the limit under the differentiation sign.*

Let x be variable in the interval (a, b) and a be any constant. We have by property (2):

$$\lim_{n\to\infty} \int_a^x s_n'(x)\,dx = \int_a^x \sigma(x)\,dx.$$

But

$$\int_a^x s_n'(x)\,dx = s_n(x) - s_n(a) \to s(x) - s(a),$$

and therefore the previous formula gives:

$$s(x) - s(a) = \int_a^x \sigma(x)\,dx.$$

Differentiating this equation and using a familiar property of definite integrals (property VII) [95], we have:

$$\frac{ds(x)}{dx} = \sigma(x),$$

which it was required to prove. It remains to prove the uniform convergence of sequence $s_n(x)$. We have:

$$s_n(x) = s_n(a) + \int_a^x s_n'(x)\,dx.$$

Sequence $s_n(a)$ is convergent, and it does not contain x. Sequence $\int_a^x s_n'(x)\,dx$ is uniformly convergent by property 2. Hence follows the uniform convergence of $s_n(x)$, since it is a direct result of the definition of uniform convergence that the sum of two uniformly convergent sequences is also a uniformly convergent sequence. Of course, every convergent sequence, the terms of which do not contain x, as for instance, $s_n(a)$, falls under the definition of uniformly convergent sequence.

We also notice that use has only been made of the uniform convergence of $s_n'(x)$ and of the convergence of $s_n(a)$ in order to prove the uniform convergence of $s_n(x)$ throughout (a, b); it is thus sufficient to demand the convergence of $s_n(x)$ at a single point $x = a$ when formulating the last property. The uniform convergence of $s_n(x)$ throughout (a, b) follows from this, as we have already said.

146. Properties of uniformly convergent series. If we take $s_n(x)$ in the above propositions as the sum of the first n terms of a given series

$$u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots,$$

and $s(x)$ as the sum of the whole series, we at once obtain analogous propositions for series with variable terms.

1. *If the terms of the series*

$$u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots \tag{55}$$

are continuous in the interval (a, b), and the series is uniformly convergent, its sum $s(x)$ is also continuous in (a, b).

2. *If the terms of series (55) are continuous functions in the interval (a, b) and the series is uniformly convergent, it can be integrated term by term between any desired limits α, β, lying in (a, b),* i.e.

$$\int_\alpha^\beta \sum_{n=1}^{\infty} u_n(x)\,dx = \sum_{n=1}^{\infty} \int_\alpha^\beta u_n(x)\,dx. \tag{56}$$

If the limits of integration are variable, e.g. $\beta = x$, *the series obtained by term-by-term integration :*

$$\int_a^x u_1(x)\,\mathrm{d}x + \int_a^x u_2(x)\,\mathrm{d}x + \ldots + \int_a^x u_n(x)\,\mathrm{d}x + \ldots, \tag{57}$$

is also uniformly convergent in the interval (a, b).

3. *If the series* (55) *is convergent in the interval* (a, b) *and the derivatives of its terms,* $u_1'(x), u_2'(x), \ldots, u_n'(x), \ldots$, *are continuous in* (a, b), *the series formed by the derivatives*

$$u_1'(x) + \ldots + u_n'(x) + \ldots$$

being uniformly convergent in (a, b), *the given series* (55) *is uniformly convergent, and can be differentiated term by term,* i.e.

$$\frac{\mathrm{d}}{\mathrm{d}x} \sum_{n=1}^{\infty} u_n(x) = \sum_{n=1}^{\infty} \frac{\mathrm{d}u_n(x)}{\mathrm{d}x}. \tag{58}$$

It need only be borne in mind, in order to deduce these propositions from the theorem of [145], that, as we know already, *the properties indicated in the propositions are valid in the case of a finite number of terms.* For instance, *if the terms* $u_n(x)$ *of the series are continuous, the function*

$$s_n(x) = u_1(x) + u_2(x) + \ldots + u_n(x)$$

is continuous for any n [34].

147. Tests for uniform convergence. We state sufficient conditions for uniform convergence. *A series of functions, defined in an interval* (a, b):

$$u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots$$

is uniformly convergent in (a, b) *if one of the following conditions is satisfied*:

(A) *A sequence of positive constants can be found,*

$$M_1, M_2, \ldots, M_n, \ldots$$

such that

$$| u_n(x) | \leqslant M_n \quad \text{in} \quad (a, b) \tag{59}$$

and the series

$$M_1 + M_2 + \ldots + M_n + \ldots \tag{60}$$

is convergent (Weierstrass's test).

(B) *The functions* $u_n(x)$ *can be put in the form :*

$$u_n(x) = a_n v_n(x)\,, \tag{61}$$

where $a_1, a_2, \ldots, a_n, \ldots$ *are constants, such that the series*

$$a_1 + a_2 + \ldots + a_n + \ldots \tag{62}$$

is convergent, and functions $v_1(x), \ldots, v_n(x), \ldots$ *are all non-negative and less than a positive constant* M, *and satisfy for all* x *of* (a, b) *:*

$$v_1(x) \geqslant v_2(x) \geqslant \ldots \geqslant v_n(x) \geqslant \ldots;\ v_n(x) \leqslant M \tag{63}$$

(Abel's test).

Proof (A). Since series (60) is convergent, given ε, an N can be found such that, for all $n > N$ and for all p [125]:

$$M_{n+1} + M_{n+2} + \dots + M_{n+p} < \varepsilon\,;$$

and using (59):

$$|\,u_{n+1}(x) + \dots + u_{n+p}(x)\,| \leqslant M_{n+1} + \dots + M_{n+p} < \varepsilon\,,$$

whence follows [143] the uniform convergence of series (55).

Proof (B). We put:

$$\sigma'_p = a_{n+1} + a_{n+2} + \dots + a_{n+p} \qquad (p = 1, 2, \dots)\,,$$

which gives us at once:

$$a_{n+1} = \sigma'_1 \quad \text{and} \quad a_{n+k} = \sigma'_k - \sigma'_{k-1} \qquad (k > 1)\,.$$

We find:

$$u_{n+1}(x) + u_{n+2}(x) + \dots + u_{n+p}(x) =$$
$$= a_{n+1}v_{n+1}(x) + a_{n+2}v_{n+2}(x) + \dots + a_{n+p}v_{n+p}(x)\,.$$

We substitute here for a_{n+k} in terms of σ'_k, and collect terms in the same σ'_k, giving:

$$\begin{aligned} &a_{n+1}v_{n+1}(x) + a_{n+2}v_{n+2}(x) + \dots + a_{n+p}v_{n+p}(x) = \\ &= \sigma'_1 v_{n+1}(x) + (\sigma'_2 - \sigma'_1)\,v_{n+2}(x) + \dots + (\sigma'_p - \sigma'_{p-1})\,v_{n+p}(x) = \\ &= \sigma'_1\left[v_{n+1}(x) - v_{n+2}(x)\right] + \dots + \\ &\qquad + \sigma'_{p-1}\left[v_{n+p-1}(x) - v_{n+p}(x)\right] + \sigma'_p v_{n+p}(x)\,. \end{aligned}$$

Recalling that $v_{n+p}(x)$ and all the differences $v_{n+k-1}(x) - v_{n+k}(x)$ are non-negative by hypothesis, we can write:

$$\begin{aligned} |\,u_{n+1}(x) + \dots + u_{n+p}(x)\,| &\leqslant |\,\sigma'_1\,|\left[v_{n+1}(x) - v_{n+2}(x)\right] + \dots + \\ &+ |\,\sigma'_{p-1}\,|\left[v_{n+p-1}(x) - v_{n+p}(x)\right] + |\,\sigma'_p\,|\,v_{n+p}(x)\,, \end{aligned}$$

or, letting σ' denote the greatest of the absolute values $|\,\sigma'_1\,|, |\,\sigma'_2\,|, \dots, |\,\sigma'_p|$,

$$\begin{aligned} &|\,u_{n+1}(x) + \dots + u_{n+p}(x)\,| \leqslant \\ &\leqslant \sigma'\left\{[v_{n+1}(x) - v_{n+2}(x)] + \dots + [v_{n+p-1}(x) - v_{n+p}(x)] + v_{n+p}(x)\right\} \end{aligned}$$

which gives after cancelling:

$$|\,u_{n+1}(x) + \dots + u_{n+p}(x)\,| \leqslant \sigma' v_{n+1}(x)\,. \tag{64}$$

It follows from the definition of σ'_k and the convergence of series (62) that, for any given positive ε, there exists an N such that

$$|\,\sigma'_k\,| < \frac{\varepsilon}{M}$$

for $n > N$ and any k; and hence:

$$\sigma' < \frac{\varepsilon}{M}\,.$$

Bearing in mind that also, $0 \leqslant v_{n+p}(x) \leqslant M$, by hypothesis, we get from (64):

$$|u_{n+1}(x) + \ldots + u_{n+p}(x)| < \varepsilon$$

for $n > N$ and any p. Since N does not depend on x, the uniform convergence of series (55) in (a, b) follows from this.

Examples.

1. The series

$$\sum_{n=1}^{\infty} \frac{\cos nx}{n^p}, \quad \sum_{n=1}^{\infty} \frac{\sin nx}{n^p} \qquad (p > 1) \tag{65}$$

are uniformly convergent in any interval, since we have for any x:

$$\left| \frac{\cos nx}{n^p} \right| \leqslant \frac{1}{n^p}, \quad \left| \frac{\sin nx}{n^p} \right| \leqslant \frac{1}{n^p},$$

and the series $\sum \frac{1}{n^p}$ is convergent for $p > 1$ [122] (Weierstrass's test).

2. If $\sum\limits_{n=1}^{\infty} a_n$ is convergent, the series

$$\sum_{n=1}^{\infty} \frac{a_n}{n^x} \tag{66}$$

is uniformly convergent in the interval $(0 \leqslant x \leqslant l)$ for any l, since, on substituting here

$$v_n(x) = \frac{1}{n^x},$$

all the conditions of Abel's test are satisfied.

148. Power series. Radius of convergence. One of the most important applications of the above theory of series with variable terms is the power series, i.e. the series of the form:

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots, \tag{67}$$

already encountered by us in the discussion of Maclaurin's formula. The detailed study of the properties of these series belongs to the theory of functions of a complex variable, so that we only indicate here the most fundamental properties.

ABEL'S FIRST THEOREM. *If the power series* (67) *is convergent for a certain value of* $x = \xi$, *it is absolutely convergent for all* x *satisfying*

$$|x| < |\xi|. \tag{68}$$

Conversely, *if it is divergent for* $x = \xi$, *it is also divergent for all* x *satisfying*

$$|x| > |\xi| = r. \tag{69}$$

Firstly, let the series

$$a_0 + a_1\xi + a_2\xi^2 + \ldots + a_n\xi^n + \ldots$$

be convergent. Since the general term of a convergent series must tend to zero,

$$a_n\xi^n \to 0 \quad \text{as} \quad n \to \infty ,$$

and hence a constant M can be found, such that for all n:

$$| a_n\xi^n | < M .$$

We now assign to x any value satisfying (68), and put

$$q = \left| \frac{x}{\xi} \right| < 1 .$$

We obviously have:

$$| a_n x^n | = \left| a_n \xi^n \frac{x^n}{\xi^n} \right| = | a_n \xi^n | \left| \frac{x}{\xi} \right|^n < Mq^n ,$$

i.e. the absolute value of the general term of (67), for the x in question, does not exceed the general term of an infinite decreasing geometrical progression. Series (67) is therefore absolutely convergent [124].

The second part of the theorem is obvious, since if (67) were convergent for some x satisfying (69), by what has just been proved, it would have to be convergent for any ξ satisfying $| \xi | < | x |$, which contradicts what is given.

Corollary. *There exists a fully defined number R, called the radius of convergence of series* (67), *and having the following properties*:

series (67) *is absolutely convergent for* $| x | < R$,
,, (67) *is divergent for* $| x | > R$.

In particular, it may happen that $R = 0$, in which case series (67) is divergent for all x differing from zero; or that $R = \infty$, in which case (67) is convergent for all x.

We dismiss the first case, and consider a positive value of $x = \xi$, for which (67) is convergent. Such a value certainly exists, if there exist in general values of $x \neq 0$ for which (67) is convergent. If we increase the number ξ, only two cases can arise: either (67) always remains convergent for $x = \xi$, even when ξ increases indefinitely, in which case we obviously have $R = \infty$; or else there exists some constant A, with the property that, however close ξ approaches A, whilst remaining less than A, series (67) is always convergent, yet when ξ is greater than A, the series becomes divergent.

The existence of such an A is quite obvious on intuitive-geometric grounds, since, on the basis of Abel's first theorem, if the series is divergent for any given value ξ, it will also be divergent for all greater values. The existence of A can be proved rigorously from the theory of irrational numbers. This number A is obviously the radius of convergence R of series (67).

We give the proof of the existence of R. We divide all real numbers into two classes as follows: we put in the first class all negative numbers, zero, and the positive ξ for which (67) is convergent with $| x | = \xi$, and we put in the second class all the remaining real numbers. By the theorem proved, any number of the first class is less than any number of the second class, i.e. we have made a section in the domain of real numbers. There is therefore

either a greatest number in the first class, or a least number in the second class [40]. It is easily seen that this number is the radius of convergence R of the series. If every number falls into the first class, we have to take $R = \infty$.

149. Abel's second theorem. *If R is the radius of convergence of series* (67), *the series is uniformly, as well as absolutely, convergent in any interval* (a, b), *lying wholly within the interval* $(-R, +R)$, i.e. *for which*

$$-R < a < b < R.$$

If the series is also convergent for $x = R$ or for $x = -R$, it will be uniformly convergent in (a, R) *or* $(-R, b)$.

We note, first of all, that we can take $R = 1$ without loss of generality, by taking a new independent variable t instead of x, defined by

$$x = Rt.$$

Series (67) now becomes a power series in t, and the interval $(-R, +R)$ becomes $(-1, 1)$.

If $R = 1$, it follows from the definition of radius of convergence that (67) is absolutely convergent for any $x = \xi$, where $|\xi| < 1$. We now take any interval (a, b) inside $(-R, R)$, so that

$$-1 < a < b < 1.$$

We choose for ξ any number lying inside $(-1, 1)$, but greater in absolute value than $|a|$ and $|b|$. We have for every x in (a, b):

$$|a_n x^n| < |a_n \xi^n|,$$

and since the series

$$a_0 + a_1\xi + a_2\xi^2 + \ldots + a_n\xi^n + \ldots$$

is absolutely convergent and its terms are independent of x, it follows by Weierstrass's test that (67) is uniformly convergent in (a, b).

We now suppose that (67) is also convergent for $x = 1$, i.e. that

$$a_0 + a_1 + a_2 + \ldots + a_n + \ldots$$

is convergent. Putting

$$v_n(x) = x^n,$$

we can apply Abel's test to series (67), showing that (67) is uniformly convergent throughout an interval $(a, 1)$, where a is any number greater than -1.

The case of (67) converging for $x = -1$ follows from the above on substituting $(-x)$ for x.

Let $f(x)$ denote the sum of series (67). It only exists, of course, for the x for which the series converges. Let R be the radius of convergence of the series. Bearing in mind the uniform convergence of the series in any interval (a, b), for which

$$-R < a < b < R, \tag{70}$$

and also property 1 of [146], we can assert that *the sum* $f(x)$ *of the series is continuous in any of the above intervals* (a, b). In other words, $f(x)$ *is continuous inside* $(-R, +R)$. We also see that this function has every type of derivative inside $(-R, +R)$. If series (67) also converges for $x = R$, it follows from the uniform convergence proved for any interval (a, R), where $a > -R$, that $f(x)$ is continuous in this interval, and in particular, $f(R)$ is the limit of $f(x)$ on x tending to R from the left [35]:

$$f(R) = \lim_{x \to R-0} f(x). \tag{71}$$

Similarly for the convergence of the series for $x = -R$.

We saw above that Newton's binomial expansion [131]:

$$(1 + x)^m = 1 + \frac{m}{1!} x + \frac{m(m-1)}{2!} x^2 + \ldots$$

has a radius of convergence $R = 1$, and converges in certain cases for $x = \pm 1$. We are able to say, by what has just been proved, that if the series converges for $x = 1$, for instance, its sum will then be:

$$\lim_{x \to 1-0} (1 + x)^m = 2^m.$$

150. Differentiation and integration of power series. Let R be the radius of convergence of the series

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots \tag{72}$$

We get two further power series by integrating (72) term by term from 0 to x, and by differentiating it:

$$a_0 x + \frac{a_1}{2} x^2 + \ldots + \frac{a_n}{n+1} x^{n+1} + \ldots \tag{73}$$

$$a_1 + 2a_2 x + 3a_3 x^2 + \ldots + n a_n x^{n-1} + \ldots \tag{74}$$

We show that these have the same radius of convergence R. We have to show, for this, that they are convergent if $|x| < R$.

Series (72) is uniformly convergent in any interval $(-R_1, +R_1)$, where $0 < R_1 < R$, by what has been proved; and by property 2 of [146] it can be integrated term by term from 0 to x in this interval, i.e. we can say that series (73) is convergent for any x, for which $|x| < R$, and that the sum of (73) is then

$$\int_0^x f(x)\, dx,$$

where $f(x)$ is the sum of series (72). We now show that series (74) is also convergent if $|x| < R$. We take such an x, then choose a ξ, lying between $|x|$ and R, i.e.

$$|x| < \xi < R, \tag{75}$$

and put

$$q = \frac{|x|}{\xi} < 1.$$

We now have for the terms of series (74):

$$|na_n x^{n-1}| = \left| na_n \xi^n \frac{x^{n-1}}{\xi^{n-1}} \cdot \frac{1}{\xi} \right|,$$

and by the above:

$$|na_n x^{n-1}| < nq^{n-1} \frac{1}{\xi} |a_n \xi^n|.$$

It is easy to show, by applying d'Alembert's test to the series Σnq^{n-1}, that it is convergent for $0 < q < 1$, and hence [119]:

$$nq^{n-1} \to 0 \quad \text{as} \quad n \to \infty, \tag{76}$$

so that, for all sufficiently large n:

$$|na_n x^{n-1}| < |a_n \xi^n|.$$

But by (75), the series $\Sigma a_n \xi^n$ is absolutely convergent, and therefore series (74) is absolutely convergent for the x taken. Both series (73) and (74) are thus convergent if $|x| < R$, i.e. the radius of convergence of a power series is not decreased by term by term integration or differentiation. But it immediately follows from this that it can also not increase. Suppose, for instance, that the radius of convergence of series (73) were R', where $R' > R$, then differentiation of series (73) would give us series (72), and its radius of convergence would not be less than R', by what has just been proved; but in fact it is R, where $R < R'$. Thus, series (73) and (74) have the same radius of convergence R as series (72). Differentiating (74) a second time gives us a power series

$$2a_2 + 3 \cdot 2a_3 x + 4 \cdot 3a_4 x^2 + \ldots + n(n-1) a_n x^{n-2} + \ldots$$

with the same radius of convergence R, by what has been proved; and so on. All these power series are uniformly convergent in any interval (a, b) for which (70) is true; and the same applies for repeated term by term integration. Recalling properties 2 and 3 of [146], we can finally state the following result:

The power series

$$a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots, \tag{77}$$

the radius of convergence of which is equal to R, is a continuous function of x inside the interval $(-R, R)$.

The series may be differentiated and integrated term by term any number of times, if x lies inside $(-R, R)$, the power series obtained in this way having the same radius of convergence R. The sums of these latter series are equal to the corresponding derivatives and integrals of the sum of series (77).

Putting

$$f(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_n x^n + \ldots, \tag{78}$$

we obtain in this way:

$$f'(x) = a_1 + 2a_2x + \ldots + na_nx^{n+1} + \ldots ,$$
$$f''(x) = 2a_2 + 6a_3x + \ldots + n(n-1)\, a_nx^{n-2} + \ldots ,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$f^{(n)}(x) = n!\, a_n + (n+1)\, n \ldots 3 \cdot 2a_{n+1}x + \ldots ,$$

which give with $x = 0$:

$$a_0 = f(0)\,, \quad a_1 = \frac{f'(0)}{1!}\,, \quad a_2 = \frac{f''(0)}{2!}\,, \ldots, \quad a_n = \frac{f^{(n)}(0)}{n!}\,, \ldots$$

Substituting these expressions for $a_0, a_1, a_2, \ldots, a_n$ in (78), we obtain:

$$f(x) = f(0) + \frac{xf'(0)}{1!} + \frac{x^2f''(0)}{2!} + \ldots + \frac{x^n f^{(n)}(0)}{n!} + \ldots \, (-R < x < +R)\ ,$$

i.e. *a power series coincides with the expansion of its sum by Maclaurin's formula.*

The theory of power series given above is easily extended to power series of the form:

$$a_0 + a_1(x-a) + a_2(x-a)^2 + \ldots + a_n(x-a)^n + \ldots \qquad (79)$$

The role of x is played everywhere by $(x - a)$. The radius of convergence R of series (79) is defined by the condition that the series converges for $|x - a| < R$, and diverges for $|x - a| > R$. If $f(x)$ denotes the sum of (79) in the interval

$$-R < x - a < R\,, \qquad (80)$$

we obtain the expressions for the coefficients a_n:

$$a_0 = f(a)\,, \quad a_1 = \frac{f'(a)}{1!}, \quad \ldots, a_n = \frac{f^{(n)}(a)}{n!}\,, \ldots,$$

i.e. *series* (79) *coincides in the interval* (80) *with the expansion of its sum into a Taylor series.*

We return to the theory of power series in the third volume, when dealing with the theory of functions of a complex variable.

We propose as examples that the expansions of $\log(1+x)$, arc tan x, and arc sin x be deduced from the theory of power series, noticing that

$$\log(1+x) = \int_0^x \frac{dx}{1+x}\,,$$

$$\text{arc}\tan x = \int_0^x \frac{dx}{1+x^2}\,,$$

$$\text{arc}\sin x = \int_0^x \frac{dx}{\sqrt{1-x^2}}\,;$$

the domains of application of the expansions obtained should also be investigated.

EXERCISES ON CHAPTER IV

Write down the formula for the nth term of each of the series **1–10**:

1. $1+\frac{1}{3}+\frac{1}{5}+\frac{1}{7}+\ldots$.

2. $\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\frac{1}{8}+\ldots$.

3. $1+\frac{2}{2}+\frac{3}{4}+\frac{4}{8}+\ldots$.

4. $1+\frac{1}{4}+\frac{1}{9}+\frac{1}{16}+\ldots$.

5. $\frac{3}{4}+\frac{4}{9}+\frac{5}{16}+\frac{6}{25}+\ldots$.

6. $\frac{2}{5}+\frac{4}{8}+\frac{6}{11}+\frac{8}{14}+\ldots$.

7. $\frac{1}{2}+\frac{1}{6}+\frac{1}{12}+\frac{1}{20}+\frac{1}{30}+\frac{1}{42}+\ldots$.

8. $1+\frac{1\cdot 3}{1\cdot 4}+\frac{1\cdot 3\cdot 5}{1\cdot 4\cdot 7}+\frac{1\cdot 3\cdot 5\cdot 7}{1\cdot 4\cdot 7\cdot 10}+\ldots$.

9. $1-1+1-1+1-1+\ldots$.

10. $1+\frac{1}{2}+3+\frac{1}{4}+5+\frac{1}{6}+\ldots$.

Examine for convergence the series **11–21** using the comparison test:

11. $1-1+1-1+\ldots+(-1)^{n-1}+\ldots$.

12. $\frac{2}{5}+\frac{1}{2}\left(\frac{2}{5}\right)^2+\frac{1}{3}\left(\frac{2}{5}\right)^3+\ldots+\frac{1}{n}\left(\frac{2}{5}\right)^n+\ldots$

13. $\frac{2}{3}+\frac{3}{5}+\frac{4}{7}+\ldots+\frac{n+1}{2n+1}+\ldots$

14. $\frac{1}{\sqrt{10}}-\frac{1}{\sqrt[3]{10}}+\frac{1}{\sqrt[4]{10}}-\ldots+\frac{(-1)^{n+1}}{\sqrt[n+1]{10}}+\ldots$.

15. $\frac{1}{2}+\frac{1}{4}+\frac{1}{6}+\ldots+\frac{1}{2n}+\ldots$.

16. $\frac{1}{11}+\frac{1}{21}+\frac{1}{31}+\ldots+\frac{1}{10n+1}+\ldots$.

17. $\frac{1}{\sqrt{1\cdot 2}}+\frac{1}{\sqrt{2\cdot 3}}+\frac{1}{\sqrt{3\cdot 4}}+\ldots+\frac{1}{\sqrt{n(n+1)}}+\ldots$.

18. $2+\frac{2^2}{2}+\frac{2^3}{3}+\ldots+\frac{2^n}{n}+\ldots$.

19. $1+\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}+\ldots+\frac{1}{\sqrt{n}}+\ldots$.

20. $\frac{1}{2^2}+\frac{1}{5^2}+\frac{1}{8^2}+\ldots+\frac{1}{(3n-1)^2}+\ldots$.

21. $\frac{1}{2}+\frac{\sqrt[3]{2}}{3\sqrt{2}}+\frac{\sqrt[3]{3}}{4\sqrt{3}}+\ldots\frac{\sqrt[3]{n}}{(n+1)\sqrt{n}}+\ldots$

With the help of d'Alembert's test investigate the convergence of the series **22–23**:

22. $\frac{1}{\sqrt{2}}+\frac{3}{2}+\frac{5}{2\sqrt{2}}+\ldots+\frac{2n-2}{(\sqrt{2})^n}+\ldots$

23. $\frac{2}{1}+\frac{2\cdot5}{1\cdot5}+\frac{2\cdot5\cdot8}{1\cdot5\cdot9}+\ldots+\frac{2\cdot5\cdot8\ldots(3n-1)}{1\cdot5\cdot9\ldots(4n-3)}+\ldots$

With the help of Cauchy's test investigate the convergence of the series **24–25**:

24. $\frac{2}{1}+\left(\frac{3}{3}\right)^2+\left(\frac{4}{5}\right)^3+\ldots+\left(\frac{n+1}{2n-1}\right)^n+\ldots$

25. $\frac{1}{2}+\left(\frac{2}{5}\right)^3+\left(\frac{3}{8}\right)^5+\ldots+\left(\frac{n}{3n-1}\right)^{2n-1}+\ldots.$

Investigate the convergence of the series **26–54**:

26. $1+\frac{1}{2!}+\frac{1}{3!}+\ldots+\frac{1}{n!}+\ldots.$

27. $\frac{1}{3}+\frac{1}{8}+\frac{1}{15}+\ldots+\frac{1}{(n+1)^2-1}+\ldots.$

28. $\frac{1}{1\cdot4}+\frac{1}{4\cdot7}+\frac{1}{7\cdot10}+\ldots+\frac{1}{(3n-2)(3n+1)}+\ldots$

29. $\frac{1}{3}+\frac{4}{9}+\frac{4}{19}+\ldots+\frac{n^2}{2n^2+3}+\ldots.$

30. $\frac{1}{2}+\frac{2}{5}+\frac{3}{10}+\ldots+\frac{n}{n^2+1}+\ldots.$

31. $\frac{3}{2^2\,3^2}+\frac{5}{3^2\,4^2}+\frac{7}{4^2\,5^2}+\ldots+\frac{2n+1}{(n+1)^2\,(n+2)^2}+\ldots$

32. $\frac{3}{4}+\left(\frac{6}{7}\right)^2+\left(\frac{9}{10}\right)^3+\ldots+\left(\frac{3n}{3n+1}\right)^n+\ldots$

33. $\left(\frac{3}{4}\right)^{\frac{1}{2}}+\frac{5}{7}+\left(\frac{7}{10}\right)^{\frac{3}{2}}+\ldots+\left(\frac{2n+1}{3n+1}\right)^{\frac{1}{2}n}+\ldots$

34. $\frac{1}{e}+\frac{8}{e^2}+\frac{27}{e^3}+\ldots+\frac{n^3}{e^n}+\ldots$

35. $1+\frac{2}{2^2}+\frac{4}{3^3}+\ldots+\frac{2^{n-1}}{n^n}+\ldots$

36. $\sum_{n=1}^{\infty}\arcsin n^{-\frac{1}{2}}.$ **37.** $\sum_{n=1}^{\infty}\sin(n^{-2}).$

38. $\sum_{n=1}^{\infty}\log\left(1+\frac{1}{n}\right).$ **39.** $\sum_{n=1}^{\infty}\log\left(1+\frac{1}{n^2}\right).$

40. $\sum_{n=1}^{\infty}(\log n)^{-1}$ **41.** $\sum_{n=1}^{\infty}(n\log n)^{-1}$

42. $\sum_{n=1}^{\infty}n^{-1}(\log n)^{-2}.$ **43.** $\sum_{n=2}^{\infty}\frac{1}{n\cdot\log n\cdot\log\log n}.$

44. $\sum_{n=1}^{\infty}(n^2-n)^{-1}.$ **45.** $\sum_{n=1}^{\infty}\left[n\,(n+1)\right]^{-\frac{1}{2}}.$

46. $\sum_{n=1}^{\infty}\left[n\,(n+1)\,(n+2)\right]^{-\frac{1}{2}}.$ **47.** $\sum_{n=2}^{\infty}\frac{1}{n\log n+\log^{3\cdot2}n}.$

48. $\sum_{n=2}^{\infty}\frac{1}{n^{4/3}-n^{1/2}}.$ **49.** $\sum_{n=1}^{\infty}\frac{n^{1/3}}{(2n-1)(5n^{1/3}-1)}.$

50. $\sum_{n=1}^{\infty}\left(1-\cos\frac{\pi}{n}\right)$. **51.** $\sum_{n=1}^{\infty}\frac{n!}{n^n}$. **52.** $\sum_{n=1}^{\infty}\frac{2^n n!}{n^n}$.

53. $\sum_{n=1}^{\infty} 3^n n^{-n} n!$ **54.** $\sum_{n=1}^{\infty} e^n n^{-n} n!$

Investigate the convergence of the series **55–66.** In the case of convergent series state whether the series is absolutely or conditionally convergent.

55. $1-\frac{1}{3}+\frac{1}{5}-\ldots+\frac{(-1)^{n-1}}{2n-1}+\ldots$

56. $1-\frac{1}{\sqrt{2}}+\frac{1}{\sqrt{3}}-\ldots+\frac{(-1)^{n-1}}{\sqrt{n}}+\ldots$

57. $1-\frac{1}{4}+\frac{1}{9}-\ldots+\frac{(-1)^{n-1}}{n^2}+\ldots$

58. $1-\frac{2}{7}+\frac{3}{13}-\ldots+\frac{(-1)^{n-1}n}{6n-5}+\ldots.$

59. $\frac{3}{1\cdot 2}-\frac{5}{2\cdot 3}+\frac{7}{3\cdot 4}-\ldots+(-1)^{n-1}\frac{2n+1}{n(n+1)}+\ldots$

60. $-\frac{1}{2}-\frac{2}{4}+\frac{3}{8}+\frac{4}{16}-\ldots+(-1)^{(n^2+n)/2}\frac{n}{2^n}+\ldots$

61. $-\frac{2}{2\sqrt{2}-1}+\frac{3}{3\sqrt{3}-1}-\frac{4}{4\sqrt{4}-1}+\ldots+$
$+(-1)^n\frac{n+1}{(n+1)\sqrt{(n+1)}-1}+\ldots$

62. $-\frac{3}{4}+\left(\frac{5}{7}\right)^2-\left(\frac{7}{10}\right)^3+\ldots+(-1)^n\left(\frac{2n+1}{3n+1}\right)^n+\ldots$

63. $\frac{3}{2}-\frac{3\cdot 5}{2\cdot 5}+\frac{3\cdot 5\cdot 7}{2\cdot 5\cdot 8}-\ldots+$
$+(-1)^{n-1}\frac{3\cdot 5\cdot 7\ldots(2n+1)}{2\cdot 5.8\ldots(3n-1)}+\ldots$

64. $\frac{1}{7}-\frac{1\cdot 4}{7\cdot 9}+\frac{1\cdot 4\cdot 7}{7\cdot 9\cdot 11}-\ldots+$
$+(-1)^{n-1}\frac{1\cdot 4\cdot 7\ldots(3n-2)}{7\cdot 9\cdot 11\ldots(2n+5)}+\ldots$

65. $\sum_{n=1}^{\infty}\frac{\sin n\alpha}{(\log 10)^n}$. **66.** $\sum_{n=1}^{\infty}(-1)^n\frac{\log n}{n}$.

Investigate the convergence of the series **67–73** with complex terms:

67. $\sum_{n=1}^{\infty}\frac{n(2+i)^n}{2^n}$. **68.** $\sum_{n=1}^{\infty}\frac{n(2i-1)^n}{3^n}$.

69. $\sum_{n=1}^{\infty}\frac{1}{n(3+i)^n}$. **70.** $\sum_{n=1}^{\infty}\frac{i^n}{n}$. **71.** $\sum_{n=1}^{\infty}\frac{1}{i-\sqrt{n}}$.

72. $\sum_{n=1}^{\infty}\frac{1}{(n+i)\sqrt{n}}$. **73.** $\sum_{n=1}^{\infty}\frac{1}{[n+(2n-1)i]^2}$.

74. Construct two series with the property that their sum is convergent while their difference is divergent.

75. Construct the product of the series $\sum_{n=1}^{\infty} n^{-3/2}$, $\sum_{n=1}^{\infty} 2^{-n+1}$. Is this product convergent?

76. Construct the series $\left(\sum_{n=1}^{\infty} 2^{1-n}\right)^2$ and consider its convergence.

77. Given the series $\sum_{n=1}^{\infty} (-1)^n/n!$ Estimate the error involved in taking the sum of the series to be given by the sum of the first four terms. Estimate it also when the sum of the first five terms is taken. What is it possible to say about the signs of these errors?

78. Estimate the error in replacing the series $\sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^n \frac{1}{n!}$ by the sum of its first n terms.

79. Estimate the error in replacing the series $\sum_{n=1}^{\infty} \frac{1}{n!}$ by the sum of its first n terms. In particular estimate the error corresponding to $n = 10$.

Find the interval of convergence of the following series **80–93** and investigate the convergence at the end-points of the interval.

80. $\sum_{n=1}^{\infty} x^n$. **81.** $\sum_{n=1}^{\infty} n^{-1} 2^{-n} x^n$. **82.** $\sum_{n=1}^{\infty} x^{2n-1}/(2n-1)$.

83. $\sum_{n=1}^{\infty} (4n-3)^{-2} 2^{n-1} x^{2n-1}$. **84.** $\sum_{n=1}^{\infty} (-1)^{n-1} n^{-1} x^n$.

85. $\sum_{n=0}^{\infty} (n+1)^5 x^{2n}/(2n+1)$. **86.** $\sum_{n=0}^{\infty} (-1)^n (2n+1)^2 x^n$.

87. $\sum_{n=0}^{\infty} x^n/n!$ **88.** $\sum_{n=1}^{\infty} n!\, x^n$. **89.** $\sum_{n=1}^{\infty} n^{-n} x^n$.

90. $\sum_{n=1}^{\infty} \left(\frac{n}{2n+1}\right)^{2n-1} x^n$. **91.** $\sum_{n=1}^{\infty} 3^{n^2} x^{n^2}$.

92. $\sum_{n=1}^{\infty} \frac{n}{n+1}\left(\frac{x}{2}\right)^n$. **93.** $\sum_{n=1}^{\infty} n!\, n^{-n} x^n$.

Determine the circle of convergence for the series **94–100**:

94. $\sum_{n=0}^{\infty} i^n z^n$. **95.** $\sum_{n=0}^{\infty} (1+ni) z^n$.

96. $\sum_{n=1}^{\infty} \frac{(z-2i)^n}{n\, 3^n}$. **97.** $\sum_{n=1}^{\infty} \frac{z^{2n}}{2^n}$.

98. $(1+2i)+(1+2i)(3+2i)z+\ldots+$
$+(1+2i)(3+2i)\ldots(2n+1+2i)z^n+\ldots$

99. $1+\frac{z}{1-i}+\frac{z^2}{(1-i)(1-2i)}+\ldots+$
$+\frac{z^n}{(1-i)(1-2i)\ldots(1-ni)}+\ldots$

100. $\sum_{n=0}^{\infty}\left(\frac{1+2ni}{n+2i}\right)^n z^n.$

Prove the uniform convergence of the series **101-103** in the intervals stated:

101. $\sum_{n=1}^{\infty} n^{-2}x^n$ in the closed interval $[-1, 1]$. **102.** $\sum_{n=1}^{\infty} 2^{-n}\sin nx$ on the whole real line. **103.** $\sum_{n=1}^{\infty}(-1)^{n-1}n^{-1/2}x^n$ in the closed interval $[0, 1]$.

Applying term-by-term differentiation or integration find the sums of the series **104–110**:

104. $x+\frac{x^2}{2}+\frac{x^3}{3}+\ldots+\frac{x^n}{n}+\ldots.$

105. $x-\frac{x^2}{2}+\frac{x^3}{3}-\ldots+(-1)^{n-1}\frac{x^n}{n}+\ldots.$

106. $x+\frac{x^3}{3}+\frac{x^5}{5}+\ldots+\frac{x^{2n-1}}{2n-1}+\ldots.$

107. $x-\frac{x^3}{3}+\frac{x^5}{5}-\ldots+(-1)^{n-1}\frac{x^{2n-1}}{2n-1}+\ldots.$

108. $1+2x+3x^2+\ldots+(n+1)x^n+\ldots$

109. $1-3x^2+5x^4-\ldots+(-1)^{n-1}(2n-1)x^{2n-2}+\ldots$

110. $1\cdot 2+2\cdot 3x+3\cdot 4x^2+\ldots+n(n+1)x^{n-1}+\ldots$

Find the sums of the series **111–114**:

111. $\frac{1}{x}+\frac{2}{x^2}+\frac{3}{x^3}+\ldots+\frac{n}{x^n}+\ldots.$

112. $x+\frac{x^5}{5}+\frac{x^9}{9}+\ldots+\frac{x^{4n-3}}{4n-3}+\ldots.$

113. $1-\frac{1}{3\cdot 3}+\frac{1}{5\cdot 3^2}-\frac{1}{7\cdot 3^3}+\ldots+\frac{(-1)^{n-1}}{(2n-1)3^{n-1}}+\ldots$

114. $\frac{1}{2}+\frac{3}{2^2}+\frac{5}{2^3}+\ldots+\frac{2n-1}{2^n}+\ldots.$

115. Expand the polynomial x^3-2x^2-5x-2 in powers of $x+4$.

116. If $f(x)=5x^3-4x^2-3x+2$, write $f(x+h)$ as a series of powers of h.

117. Expand $\log x$ as a series in $x-1$.

118. Expand $1/x$ as a series in $x-1$.

119. Expand $1/x^2$ as a series in $x + 1$.

120. Expand e^x as a series of powers of $x + 2$.

121. Expand $\sqrt{x}$ as a series of powers of $x - 4$.

122. Expand $\cos x$ as a series of powers of $(x - \pi/2)$.

Expand the functions **123–126** as power series in x, stating the interval of convergence:

123. $\sin^2 x \cos^2 x$. **124.** $(1 + x)\, e^{-x}$. **125.** $(1 + e^x)^3$. **126.** $\sqrt[3]{8 + x}$.

CHAPTER V

FUNCTIONS OF SEVERAL VARIABLES

§ 15. Derivatives and differentials

151. Basic concepts. We began the study of functions of two variables in § 6 of Chapter II by explaining the basic concepts regarding these functions. We shall now discuss functions of several variables, and also deal in more detail with the concept of limit.

We shall assume that the function $f(x, y)$ is defined either in the whole plane or in a certain domain. A definite value $f(x, y)$ thus corresponds to every point (x, y) of this domain. If interior points only of the domain are considered, the domain is said to be *open*. If the domain includes its boundary, it is said to be *closed*.

Similarly, if we take a rectangular system of Cartesian coordinates OX, OY, OZ in space, we are able to talk of a point M of space with coordinates (x, y, z), instead of talking of a set of three numbers (x, y, z). We shall assume that the function $f(x, y, z)$ is defined throughout all space, or in some domain of space that may be either open or closed. The boundaries of the domain (there may be several) will be surfaces in the simplest cases. For instance, the inequalities:

$$a_1 \leqslant x \leqslant a_2;\ b_1 \leqslant y \leqslant b_2;\ c_1 \leqslant z \leqslant c_2$$

defines a closed rectangular parallelepiped, the edges of which are parallel to the coordinate axes. The inequalities:

$$a_1 < x < a_2;\ b_1 < y < b_2;\ c_1 < z < c_2$$

define an open parallelepiped. The inequality:

$$(x - a)^2 + (y - b)^2 + (z - c)^2 \leqslant r^2$$

defines a closed sphere with centre (a, b, c) and radius r. If the equality sign is omitted, leaving only the sign $<$, an open sphere is obtained. Limit and continuity are defined for functions of three variables in exactly the same way as in [67] for two variables.

Geometrical terminology is often preserved in the case of functions $f(x_1, x_2, \ldots, x_n)$ of several variables, where $n > 3$, although the space can no longer be visualized geometrically. A sequence of n real numbers $(x_1, x_2, \ldots, x_n)$ is called a point. The totality of such points forms an n-dimensional space. A domain of the space is defined by inequalities. For instance, the inequalities:

$$c_1 \leqslant x_1 \leqslant d_1;\ c_2 \leqslant x_2 \leqslant d_2;\ \ldots;\ c_n \leqslant x_n \leqslant d_n$$

define an n-dimensional parallelepiped, or, as we sometimes say, an n-dimensional interval. The inequality

$$\sum_{k=1}^{n} (x_k - a_k)^2 \leqslant r^2$$

defines an n-dimensional sphere. The set of points defined by the last inequality with a certain choice of r, or by the inequalities $|x_k - a_k| \leqslant \varrho$ $(k = 1, 2, \ldots, n)$, where ϱ is a certain positive number, is called a *neighbourhood of the point* $(a_1, a_2, \ldots, a_n)$.

If the function $f(x_1, x_2, \ldots, x_n)$ is defined at, and in a certain neighbourhood of, the point $(a_1, a_2, \ldots, a_n)$, we say that $f(x_1, x_2, \ldots, x_n)$ tends to a limit A, or that the point $M(x_1, x_2, \ldots, x_n)$ tends to the point $M_0(a_1, a_2, \ldots, a_n)$, and we write:

$$\lim_{x_k \to a_k} f(x_1, x_2, \ldots, x_n) = A \qquad \text{or} \qquad \lim_{M \to M_0} f(x_1, x_2, \ldots, x_n) = A,$$

if for any given positive number ε there exists a positive η, such that $|A - f(x_1, x_2, \ldots, x_n)| < \varepsilon$, provided only that $|a_k - x_k| \leqslant \eta$ with $k = 1, 2, \ldots, n$, the point $M(x_1, x_2, \ldots, x_n)$ being assumed distinct from $M_0(a_1, a_2, \ldots, a_n)$. The continuity of $f(x_1, x_2, \ldots, x_n)$ at the point $M_0(a_1, a_2, \ldots, a_n)$ is defined by the equation:

$$\lim_{x_k \to a_k} f(x_1, x_2, \ldots, x_n) = f(a_1, a_2, \ldots, a_n).$$

The properties indicated in [67], for functions continuous in a closed domain, still hold.

The continuity of a sum, product and quotient of continuous functions can be justified, as in [34] for functions of a single variable. In the case of a quotient, the denominator is assumed to differ from zero at the point $(a_1, a_2, \ldots, a_n)$.

152. Passing to a limit. We dwell in more detail on the concept of limit, confining ourselves to the case of a function of two variables. If the limit exists,

$$\lim_{\substack{x \to a \\ y \to b}} f(x, y) = A, \tag{1}$$

we shall say that *a limit exists with respect to both variables*. This means, as we know [67], that $f(x, y)$ tends to a limit A as the point $M(x, y)$ tends to $M_0(a, b)$ in any prescribed manner. In particular:

$$\lim_{x \to a} f(x, b) = A \qquad \text{and} \qquad \lim_{y \to b} f(a, y) = A. \tag{2}$$

Point $M(x,y)$ tends to $M_0(a,b)$ along a line parallel to OX in the first case, and along a line parallel to OY in the second case. We remark that the existence of limit (1) does not follow, however, from the existence of limits (2) and their equality. We take as an example the function $f(x, y) = xy/(x^2 + y^2)$, and we take $a = 0$ and $b = 0$. We have:

$$\lim_{x \to 0} f(x, 0) = \lim_{x \to 0} \frac{x \cdot 0}{x^2 + 0^2} = \lim_{x \to 0} 0 = 0 \qquad \text{and} \qquad \lim_{y \to 0} f(0, y) = 0,$$

whereas limit (1) does not exist in this case. In fact, putting $y/x = \tan \alpha$, we can re-write our formula as:

$$f(x, y) = \frac{xy}{x^2 + y^2} = \frac{\tan \alpha}{1 + \tan^2 \alpha} = \sin \alpha \cos \alpha \cdot \tag{3}$$

If $M(x, y)$ tends to $M(0, 0)$ along a line passing through the origin and making an angle α_0 with OX, $f(x, y)$, as given by (3), remains constant and its magnitude depends on the choice of α_0, whence it follows that limit (1) does not exist in the example considered. We note that formula (3) does not define the function at the point $M(0,0)$ itself.

Instead of the limit (1), we can consider an *iterated limit;* here, we first pass to the limit with respect to x, with y constant and differing from b, then pass to the limit with respect to y, or vice versa:

$$\lim_{x \to a} [\lim_{y \to b} f(x, y)] \qquad \text{or} \qquad \lim_{y \to b} [\lim_{x \to a} f(x, y)] \,. \tag{4}$$

It can happen that both the iterated limits exist but are different. For instance, taking the function

$$f(x, y) = \frac{x^2 - y^2 + x^3 + y^3}{x^2 + y^2},$$

it is easy to show that:

$$\lim_{x \to 0} [\lim_{y \to 0} f(x, y)] = 1; \ \lim_{y \to 0} [\lim_{x \to 0} f(x, y)] = -1.$$

The following theorem holds, however:

THEOREM. *If the limit with respect to both variables* (1) *exists, and if for every* x, *sufficiently near to, but differing from*, a, *there exists the limit*

$$\lim_{y \to b} f(x, y) = \varphi(x), \tag{5}$$

the first iterated limit (4) *exists and is equal to* A, i.e.

$$\lim_{x \to a} \varphi(x) = A. \tag{6}$$

It follows from the existence of limit (1) [67], that for any given positive ε there exists a positive η, such that

$$|A - f(x, y)| < \varepsilon \text{ for } |x - a| < \eta \text{ and } |y - b| < \eta\,, \tag{7}$$

(x, y) being assumed not to coincide with (a, b). We fix x, different from a, so that $|x - a| < \eta$. Taking (5) into account, and passing to the limit in inequality (7), we get:

$$|A - \varphi(x)| \leqslant \varepsilon \text{ for } |x - a| < \eta \text{ and } x \neq a\,,$$

from which, in view of the arbitrariness of ε, equation (6) follows.

Note. Similarly, if we presuppose the existence of limit (1) and that, for every y sufficiently near, and differing from, b, the limit exists:

$$\lim_{x \to a} f(x, y) = \psi(y),$$

the second iterated limit (4) exists and is equal to A, i.e.

$$\lim_{y \to b} \psi(y) = A.$$

If limit (1) exists and is equal to $f(a, b)$, i.e. $A = f(a, b)$, $f(x, y)$ is continuous at the point (a, b), or, as we say, is continuous with respect to both variables at (a, b). Here, by (2):

$$\lim_{x \to a} f(x, b) = f(a, b); \qquad \lim_{y \to b} f(a, y) = f(a, b),$$

i.e. the function is continuous with respect to each individual variable at (a, b); this was discussed earlier [67]. Continuity with respect to both variables does not follow, on the other hand, from continuity with respect to each variable. Suppose, in fact, that a function is defined by (3) except at the origin, and that we put $f(0,0) = 0$. We have here, as mentioned above:

$$\lim_{x \to 0} f(x, 0) = 0 \qquad \text{and} \qquad \lim_{y \to 0} f(0, y) = 0,$$

i.e. the function is continuous with respect to each variable at (0,0). But it is not continuous with respect to both variables, since, as we have seen, $f(x, y)$ has no definite limit as $M(x, y)$ tends to $M_0(0,0)$.

If $f(x, y)$ has partial derivatives in a certain domain containing (x, y) as an interior point, the formula is valid, as we have shown [68]:

$$f(x + \Delta x, y + \Delta y) - f(x, y) = f_y'(x + \theta\Delta x, y + \Delta y)\Delta x +$$
$$+ f_y'(x, y + \theta_1\Delta y)\Delta y \quad (0 < \theta,\ \theta_1 < 1).$$

We assume that the partial derivatives are bounded in the domain concerned, i.e. their absolute values do not exceed some number M. The formula written gives in this case:

$$|f(x + \Delta x, y + \Delta y) - f(x, y)| \leqslant M(|\Delta x| + |\Delta y|)\,,$$

and the right-hand side of this inequality tends to zero as $\Delta x \to 0$ and $\Delta y \to 0$, whence it follows:

$$\lim_{\substack{\Delta x \to 0 \\ \Delta y \to 0}} f(x + \Delta x, y + \Delta y) = f(x, y),$$

i.e. *if $f(x, y)$ has bounded partial derivatives inside a certain domain, it is continuous in this domain.*

Given the additional condition $f(0,0) = 0$, function (3) is zero on the whole of OX and the whole of OY, and it evidently has partial derivatives equal to zero at $M_0(0,0)$. It also has partial derivatives at the remaining points

$$f'_x(x, y) = \frac{y^3 - x^2 y}{(x^2 + y^2)^2}; \quad f'_y(x, y) = \frac{x^3 - xy^2}{(x^2 + y^2)^2},$$

i.e. the function concerned has partial derivatives over all the plane. In spite of this, it does not possess continuity at the point (0,0), as we have seen. This is explained by the fact that the partial derivatives can take values as large as desired in absolute magnitude as the points (x, y) approach the origin.

153. Partial derivatives, and total differential of the first order. The concepts of the partial derivatives and the total differential of a function of two variables were introduced in [68]. These concepts can be extended to the case of functions of any number of variables. We take a function of four variables as an example:

$$w = f(x, y, z, t).$$

The partial derivative of this function with respect to x is defined as the limit

$$\lim_{h \to 0} \frac{f(x + h, y, z, t) - f(x, y, z, t)}{h},$$

if it exists; it is denoted by one of the symbols:

$$f'_x(x, y, z, t), \quad \text{or} \quad \frac{\partial f(x, y, z, t)}{\partial x}, \quad \text{or} \quad \frac{\partial w}{\partial x}.$$

The partial derivatives with respect to the other variables are similarly defined.

The total differential of a function is defined as the sum of its partial differentials:

$$dw = \frac{\partial w}{\partial x} dx + \frac{\partial w}{\partial y} dy + \frac{\partial w}{\partial z} dz + \frac{\partial w}{\partial t} dt,$$

where dx, dy, dz, dt are the differentials of the independent variables (arbitrary magnitudes, independent of x, y, z, t).

The differential is the principal part of the increment of the function:

$$\Delta w = f(x + dx, y + dy, z + dz, t + dt) - f(x, y, z, t),$$

and in fact (cf. [68]):

$$\Delta w = \mathrm{d}w + \varepsilon_1 \mathrm{d}x + \varepsilon_2 \mathrm{d}y + \varepsilon_3 \mathrm{d}z + \varepsilon_4 \mathrm{d}t\,,$$

where ε_1, ε_2, ε_3, ε_4 tend to zero if $\mathrm{d}x$, $\mathrm{d}y$, $\mathrm{d}z$, $\mathrm{d}t$ tend to zero, w being assumed to have continuous partial derivatives within a certain domain containing (x, y, z, t) as an interior point.

The rule for differentiating functions of a function can similarly be generalized. Suppose, for instance, that x, y, z are not independent variables, but functions of the independent variable t. In this case, w will depend on t both directly, and via x, y, z, and the total derivative of w with respect to t will have the form:

$$\frac{\mathrm{d}w}{\mathrm{d}t} = \frac{\partial w}{\partial t} + \frac{\partial w}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial w}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial w}{\partial z}\frac{\mathrm{d}z}{\mathrm{d}t}\,. \tag{8}$$

We do not dwell on the proof of this rule, since it consists of a word for word repetition of that discussed in [69]. If x, y, z depend on other independent variables besides t, we must write the partial derivatives $\partial x/\partial t$, $\partial y/\partial t$, $\partial z/\partial t$, instead of $\mathrm{d}x/\mathrm{d}t$, $\mathrm{d}y/\mathrm{d}t$, $\mathrm{d}z/\mathrm{d}t$ on the right-hand side of (8). Function w will depend in this case on the other independent variables as well as on t, and we must also replace $\mathrm{d}w/\mathrm{d}t$ by $\partial w/\partial t$ on the left-hand side of (8). But this latter partial derivative differs from the partial derivative $\partial w/\partial t$ appearing on the right-hand side of (8), the evaluation of which only takes account of the direct dependence of w on t; brackets are sometimes used to distinguish this latter partial derivative, so that equation (8) takes the form here:

$$\frac{\partial w}{\partial t} = \left(\frac{\partial w}{\partial t}\right) + \frac{\partial w}{\partial x}\frac{\partial x}{\partial t} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial t} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial t}\,. \tag{9}$$

We saw that, in the case of a function of a single variable, *the expression for its first differential does not depend on the choice of the independent variable* [50]. We show that *this property remains valid in the case of a function of several variables.*

We take for clarity a function of two variables:

$$z = \varphi(x, y)\,.$$

We suppose that x and y are functions of the independent variables u and v. We have, in accordance with the rule for differentiation of functions of a function:

$$\frac{\partial z}{\partial u} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial u} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial u}\,, \quad \frac{\partial z}{\partial v} = \frac{\partial z}{\partial x}\frac{\partial x}{\partial v} + \frac{\partial z}{\partial y}\frac{\partial y}{\partial v}\,.$$

The total differential of the function is by definition:

$$dz = \frac{\partial z}{\partial u} du + \frac{\partial z}{\partial v} dv.$$

We obtain on substituting the expressions for the partial derivatives:

$$dz = \frac{\partial z}{\partial x}\left(\frac{\partial x}{\partial u} du + \frac{\partial x}{\partial v} dv\right) + \frac{\partial z}{\partial y}\left(\frac{\partial y}{\partial u} du + \frac{\partial y}{\partial v} dv\right).$$

But the expressions in curved brackets are the total differentials of x and y, and we can write:

$$dz = \frac{\partial z}{\partial x} dx + \frac{\partial z}{\partial y} dy,$$

i.e. *the differential of a function of a function has the same form as it would have if the variables were independent.*

This property allows the rules for finding the differential of a sum, product and quotient to be extended to the case of functions of several variables:

$$d(u + v) = du + dv,\ d(uv) = v\,du + u\,dv,\ d\frac{u}{v} = \frac{v\,du - u\,dv}{v^2},$$

where u and v are functions of several independent variables. For instance, we can use the property proved to write:

$$d(uv) = \frac{\partial (uv)}{\partial u} du + \frac{\partial (uv)}{\partial v} dv = v\,du + u\,dv.$$

154. Euler's theorem. *A function of any number of variables is callep a homogeneous function of degree m of these variables if multiplication of all the variables by an arbitrary magnitude t results in multiplication of the function by t^m.*

We confine ourselves for the sake of clarity to the case of a function of two variables; we can say that $f(x, y)$ is called a homogeneous function of degree m if it satisfies the identity:

$$f(tx, ty) = t^m f(x, y). \tag{10}$$

Suppose, for instance, that $f(x, y)$ is the expression for a certain volume, that x and y are the lengths of certain lines, and that, apart from these lengths, only abstract numbers appear in the expression. Multiplication of x and y by t is equivalent to decreasing the scale t times, and it is clear that the number expressing the volume must now be multiplied by t^3, i.e. $f(x, y)$ will in this case be a homogeneous function of the third degree.†

† For instance, the volume of a cone is expressed in terms of the radius r of its base and its height h by the formula: $V = \frac{1}{3}\pi r^2 h$.

Differentiation of identity (10) with respect to t, using the rule for differentiation of a function of a function for the left-hand side, gives us the identity:

$$xf'_u(u, v) + yf'_v(u, v) = mt^{m-1} f(x, y) ,$$

where $u = tx$ and $v = ty$. We find on substituting $t = 1$ in this identity:

$$xf'_x(x, y) + yf'_y(x, y) = mf(x ,a) , \tag{11}$$

which is the expression for *Euler's theorem :*

The sum of the products of the partial derivatives of a homogeneous function and the corresponding variables, is equal to the function itself multiplied by its degree of homogeneity.

If $m = 0$, we get, on putting $t = 1/x$ in (10):

$$f(x, y) = f\left(1, \frac{y}{x}\right),$$

i.e. *a homogeneous function of degree zero is a function of the ratios of all the variables to one of them.* The sum of the products of the partial derivatives with the corresponding variables must be equal to zero for such a function. A homogeneous function of degree zero is often called simply *homogeneous.*

155. Partial derivatives of higher orders. The partial derivatives of a function of several variables are in turn functions of these variables, and we can in turn find their partial derivatives. We thus obtain partial derivatives of the second order of the original function, and these are again functions of the same variables; differentiation of them leads to partial derivatives of the third order of the original function, and so on. For instance, in the case of the function of two variables, $u = f(x, y)$, further differentiation of each of the partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ with respect to x and y leads to four second order derivatives, denoted by:

$$f''_{x^2}(x, y),\ f''_{xy}(x, y),\ f''_{yx}(x, y),\ f''_{y^2}(x, y)$$

or

$$\frac{\partial^2 f(x, y)}{\partial x^2}, \frac{\partial^2 f(x, y)}{\partial x\, \partial y}, \frac{\partial^2 f(x, y)}{\partial y\, \partial x}, \frac{\partial^2 f(x, y)}{\partial y^2},$$

or finally,

$$\frac{\partial^2 u}{\partial x^2}, \frac{\partial^2 u}{\partial x\, \partial y}, \frac{\partial^2 u}{\partial y\, \partial x}, \frac{\partial^2 u}{\partial y^2}.$$

The derivatives $f''_{xy}(x, y)$ and $f''_{yx}(x, y)$ only differ in the order of differentiation. In the first case, differentiation is first with respect

to x, then with respect to y, and in the reverse order in the second case. We show that these two derivatives are identical, i.e. *the order of differentiation does not affect the result.*

We take the expression:

$$\omega = f(x + h, y + k) - f(x + h, y) - f(x, y + k) + f(x, y).$$

Setting

$$\varphi(x, y) = f(x + h, y) - f(x, y),$$

we can write ω in the form:

$$\omega = [f(x + h, y + k) - f(x, y + k)] - [f(x + h, y) - f(x, y)] =$$
$$= \varphi(x, y + k) - \varphi(x, y).$$

Applying *Lagrange's formula* twice [63], we get:

$$\omega = k\, \varphi'_y(x, y + \theta_1 k) = [f'_y(x + h, y + \theta_1 k) - f'_y(x, y + \theta_1 k)] =$$
$$= khf''_{yx}(x + \theta_2 h, y + \theta_1 k).$$

The letter θ with either subscript denotes a number lying between 0 and 1. We denote by the symbol $f'_y(x + h, y + \theta_1 k)$ the partial derivative of $f(x, y)$ with respect to its second argument y after substitution of $x + h$ and $y + \theta_1 k$ for x and y respectively. An analogous notation is used for the other partial derivatives.

Similarly, putting:

$$\psi(x, y) = f(x, y + k) - f(x, y),$$

we can write:

$$\omega = [f(x + h, y + k) - f(x + h, y)] - [f(x, y + k) - f(x, y)] =$$
$$= \psi(x + h, y) - \psi(x, y) = h\, \psi'_x(x + \theta_3 h, y) =$$
$$= h[f'_x(x + \theta_3 h, y + k) - f_x(x + \theta_3 h, y)] =$$
$$= hkf''_{yx}(x + \theta_3 h, y + \theta_4 k).$$

We have on comparing these two expressions for ω:

$$hkf''_{yx}(x + \theta_2 h, y + \theta_1 k) = hkf''_{xy}(x + \theta_3 h, y + \theta_4 k)$$

or

$$f''_{yx}(x + \theta_2 h, y + \theta_1 k) = f''_{xy}(x + \theta_3 h, y + \theta_4 k).$$

Letting h and k tend to zero, we obtain on the assumption that the second order derivatives written are continuous:

$$f''_{yx}(x, y) = f''_{xy}(x, y).$$

This discussion leads to the following theorem:

THEOREM. *If $f(x, y)$ has continuous derivatives $f''_{yx}(x, y)$ and $f''_{xy}(x, y)$ inside a certain domain, these derivatives are equal at all interior points of the domain.*

We now consider two third order derivatives:

$$f'''_{x^2y}(x, y) \text{ and } f'''_{yx^2}(x, y),$$

differing only in the order of differentiation. Noting that *twice repeated* differentiation is independent of the order of differentiation, in accordance with the result proved, we can write:

$$f'''_{x^2y}(x, y) = \frac{\partial^2 f'_x(x, y)}{\partial x\, \partial y} = \frac{\partial^2 f'_x(x, y)}{\partial y\, \partial x} = f'''_{xyx}(x, y) =$$

$$= f'''_{yxx}(x, y) = f'''_{yx^2}(x, y),$$

i.e. in this case also, the result is not affected by the order of differentiation. This property can easily be generalized for derivatives of any order, and for functions of any number of variables, and we can state the general theorem: *the result of differentiation does not depend on the order in which the differentiations are carried out.*

We remark that use was made in the proof, not only of the *existence of the derivatives*, but also of *their continuity inside a certain domain.*

In future, we shall always assume the continuity of the derivatives that we discuss, and by the theorem just stated, it will only be necessary to indicate, for derivatives of higher orders, the order n of the derivative, the variables with respect to which differentiation is carried out, and the number of differentiations.

For instance, in the case of $w = f(x, y, z, t)$, the following notation is used:

$$\frac{\partial^n f(x, y, z, t)}{\partial x^\alpha\, \partial y^\beta\, \partial z^\gamma\, \partial t^\delta} \text{ or } \frac{\partial^n w}{\partial x^\alpha\, \partial y^\beta\, \partial y^\gamma\, \partial t^\delta} \quad (\alpha + \beta + \gamma + \delta = n),$$

indicating that the nth order derivative is taken, with differentiation carried out α times with respect to x, β times with respect to y, γ times with respect to z, and δ times with respect to t.

156. Differentials of higher orders. The total differential $\mathrm{d}u$ of a function of several variables is in turn a function of these variables, and we can define the total differential of this second function. We thus get the second order differential d^2u of the original function u, and again, this will be a function of the same variables and its total differential will bring us to the third order differential d^3u of the original function, and so on.

We consider in detail the function of two variables $u = f(x, y)$, where x and y will be assumed to be independent variables. By definition,

$$\mathrm{d}u = \frac{\partial f(x, y)}{\partial x}\,\mathrm{d}x + \frac{\partial f(x, y)}{\partial y}\,\mathrm{d}y. \tag{12}$$

When obtaining d^2u, we take into account the fact that the differentials $\mathrm{d}x$ and $\mathrm{d}y$ of the independent variables are to be regarded as constants, so that they can be taken outside the differentiation sign:

$$\begin{aligned}
\mathrm{d}^2u &= \mathrm{d}\left[\frac{\partial f(x, y)}{\mathrm{d}x}\,\mathrm{d}x\right] + \mathrm{d}\left[\frac{\partial f(x, y)}{\mathrm{d}y}\,\mathrm{d}y\right] = \\
&= \mathrm{d}x \cdot \mathrm{d}\,\frac{\partial f(x, y)}{\partial x} + \mathrm{d}y \cdot \mathrm{d}\,\frac{\partial f(x, y)}{\partial y} = \\
&= \mathrm{d}x \cdot \left[\frac{\partial^2 f(x, y)}{\partial x^2}\,\mathrm{d}x + \frac{\partial^2 f(x, y)}{\partial x\,\partial y}\,\mathrm{d}y\right] + \\
&+ \mathrm{d}y\left[\frac{\partial^2 f(x, y)}{\partial y\,\partial x}\,\mathrm{d}x + \frac{\partial^2 f(x, y)}{\partial y^2}\,\mathrm{d}y\right] = \\
&= \frac{\partial^2 f(x, y)}{\partial x^2}\,\mathrm{d}x^2 + 2\,\frac{\partial^2 f(x, y)}{\partial x\,\partial y}\,\mathrm{d}x\,\mathrm{d}y + \frac{\partial^2 f(x, y)}{\partial y^2}\,\mathrm{d}y^2.
\end{aligned}$$

We obtain d^3u in exactly the same way:

$$\begin{aligned}
\mathrm{d}^3u = \frac{\partial^3 f(x, y)}{\partial x^3}\,\mathrm{d}x^3 &+ 3\,\frac{\partial^3 f(x, y)}{\partial x^2\,\partial y}\,\mathrm{d}x^2\,\mathrm{d}y + \\
&+ 3\,\frac{\partial^3 f(x, y)}{\partial x\,\partial y^2}\,\mathrm{d}x\,\mathrm{d}y^2 + \frac{\partial^3 f(x, y)}{\partial y^3}\,\mathrm{d}y^3.
\end{aligned}$$

These expressions for d^2u and d^3u lead us to the following *symbolic formula for the differential of any order* :

$$\mathrm{d}^n u = \left(\frac{\partial}{\partial x}\,\mathrm{d}x + \frac{\partial}{\partial y}\,\mathrm{d}y\right)^n f. \tag{13}$$

The formula implies that the sum in brackets is to be worked out by Newton's binomial formula for degree n, then the various exponents of the powers of $\partial/\partial x$ and $\partial/\partial y$ are to be taken as indicating the orders of the derivatives of the function f with respect to x and y.

We have seen that (13) is justified for n equal to 1, 2, and 3. The full proof requires the use of the usual method, of assuming that the formula is true for n and proving that it is then true for $(n+1)$. Thus we find the differential of order $(n+1)$:

$$\mathrm{d}^{n+1}u = \mathrm{d}(\mathrm{d}^n u) = \frac{\partial(\mathrm{d}^n u)}{\partial x}\mathrm{d}x + \frac{\partial(\mathrm{d}^n u)}{\partial y}\mathrm{d}y = \left(\frac{\partial}{\partial x}\mathrm{d}x + \frac{\partial}{\partial y}\mathrm{d}y\right)\mathrm{d}^n u,$$

where the symbol of the general form

$$\left(\frac{\partial}{\partial x}\mathrm{d}x + \frac{\partial}{\partial y}\mathrm{d}y\right)\varphi$$

denotes

$$\frac{\partial\varphi}{\partial x}\mathrm{d}x + \frac{\partial\varphi}{\partial y}\mathrm{d}y.$$

Noting that formula (13) for $\mathrm{d}^n u$ is assumed proved, we can write:

$$\mathrm{d}^{n+1}u = \left(\frac{\partial}{\partial x}\mathrm{d}x + \frac{\partial}{\partial y}\mathrm{d}y\right)\left[\left(\frac{\partial}{\partial x}\mathrm{d}x + \frac{\partial}{\partial y}\mathrm{d}y\right)^n f\right] =$$
$$= \left(\frac{\partial}{\partial x}\mathrm{d}x + \frac{\partial}{\partial y}\mathrm{d}y\right)^{n+1} f,$$

i.e. the formula is proved for $\mathrm{d}^{n+1}u$.

Formula (13) is easily generalized for the case of a function of any number of independent variables. As we know [153], (13) is true, not merely when x and y are independent variables. When deducing the expression for $\mathrm{d}^2 u$, however, it was essential to assume that $\mathrm{d}x$ and $\mathrm{d}y$ are constants, and (13) is only true when $\mathrm{d}x$ and $\mathrm{d}y$ can be regarded as constants.

This is the case when x and y are in fact independent variables. We now suppose that x and y are linear functions of the independent variables z and t:

$$x = az + bt + c,\ y = a_1 z + b_1 t + c_1,$$

where the coefficients and the absolute terms are constants. We obtain for $\mathrm{d}x$ and $\mathrm{d}y$:

$$\mathrm{d}x = a\mathrm{d}z + b\mathrm{d}t,\ \mathrm{d}y = a_1\mathrm{d}z + b_1\mathrm{d}t.$$

But $\mathrm{d}z$ and $\mathrm{d}t$ are differentials of independent variables, and must be considered as constants; the same can therefore be said in this case as regards $\mathrm{d}x$ and $\mathrm{d}y$. We can thus assert that *symbolic formula* (13) *is valid both when x and y are independent variables, and when they are linear functions (integral polynomials of the first order) of the independent variables.*

If $\mathrm{d}x$ and $\mathrm{d}y$ cannot be regarded as constants, (13) is no longer true. We find the expression for d^2u in this general case. We are no longer justified in taking $\mathrm{d}x$ and $\mathrm{d}y$ outside the differentiation sign, as was done above, when finding

$$\mathrm{d}\left(\frac{\partial f(x,y)}{\partial x}\mathrm{d}x\right) \quad \text{and} \quad \mathrm{d}\left(\frac{\partial f(x,y)}{\partial y}\mathrm{d}y\right).$$

and the formula for differentiating a product has to be used [153].

We thus get:

$$\mathrm{d}^2u = \mathrm{d}x\mathrm{d}\frac{\partial f(x,y)}{\partial x} + \mathrm{d}y\mathrm{d}\frac{\partial f(x,y)}{\partial y} + \frac{\partial f(x,y)}{\partial x}\mathrm{d}^2x + \frac{\partial f(x,y)}{\partial y}\mathrm{d}^2y.$$

The sum of the first two terms on the right-hand side of this equation gives us the expression that we had above for d^2u, and we finally get:

$$\mathrm{d}^2u = \frac{\partial^2 f(x,y)}{\partial x^2}\mathrm{d}x^2 + 2\frac{\partial^2 f(x,y)}{\partial x\,\partial y}\mathrm{d}x\,\mathrm{d}y + \frac{\partial^2 f(x,y)}{\partial y^2}\mathrm{d}y^2 + \\ + \frac{\partial f(x,y)}{\partial x}\mathrm{d}^2x + \frac{\partial f(x,y)}{\partial y}\mathrm{d}^2y, \tag{14}$$

i.e. in the general case, the expression for d^2u contains additional terms, depending on d^2x and d^2y.

157. Implicit functions. We now indicate a rule for differentiating functions that are given implicitly. We shall assume here, that the equations written down in fact define a certain function, having the corresponding derivatives. We prove this for certain conditions in [159]. If y is an implicit function of x:

$$F(x,y) = 0. \tag{15}$$

the first derivative y' of this function is defined, as we know, by the equation [69]:

$$F'_x(x,y) + F'_y(x,y)\,y' = 0. \tag{16}$$

We derived (16) by assuming that y is a function of x in (15) and differentiating both sides of (15) with respect to x. By dealing with (16) in the same way, we get an equation for the second derivative y'':

$$F''_{x^2}(x,y) + 2F''_{xy}(x,y)\,y' + F''_{y^2}(x,y)\,y'^2 + F'_y(x,y)\,y'' = 0. \tag{17}$$

On differentiating this equation with respect to x, we get an equation for the third derivative y'', and so on.

We note the fact that, in the equations thus obtained, the coefficients of the required derivatives of the implicit function are all the same, viz., $F'_y(x, y)$, and hence, if this coefficient differs from zero for certain values of x and y satisfying (15), the above method will give, for these values, completely defined values for the derivatives of any order of the implicit function. The existence of the partial derivatives on the left-hand side of (15) is assumed here, of course.

We take the equation with three variables:

$$\Phi(x, y, z) = 0.$$

This type of equation defines z as an implicit function of the independent variables x and y, and if z were to be replaced on the left-hand side of this equation by the function of x and y that it represents, the left-hand side would become identically equal to zero. Thus, assuming that z is a function of the independent variables x and y, we must get zero on differentiating the left-hand side of the equation with respect to x, and with respect to y:

$$\Phi'_x(x, y, z) + \Phi'_z(x, y, z)\, z'_x = 0\,,$$

$$\Phi'_y(x, y, z) + \Phi'_z(x, y, z)\, z'_y = 0.$$

The partial derivatives of the first order z'_x and z'_y are found from these equations. If we differentiate the first of the relationships written a second time with respect to x, we obtain an equation for the partial derivative z''_{x^2}, and so on. The coefficient of the required derivative will be $\Phi'_z(x, y, z)$ in all the equations obtained. We now consider the system of equations:

$$\varphi(x, y, z) = 0, \;\; \psi(x, y, z) = 0.$$

We shall assume that this system defines y and z as implicit functions of x. Differentiation of both equations of the system with respect to x, on the assumption that y and z are functions of x, gives us a system of equations of the first degree, defining the derivatives y' and z' of y and z with respect to x:

$$\varphi'_x(x, y, z) + \varphi'_y(x, y, z) \cdot y' + \varphi'_z(x, y, z) \cdot z' = 0\,,$$

$$\psi'_x(x, y, z) + \psi'_y(x, y, z) \cdot y' + \psi'_z(x, y, z) \cdot z' = 0.$$

Differentiation of these relationships once more with respect to x gives us a system of equations defining the second derivatives y''

and z''. Further differentiation with respect to x gives a system of equations defining y''' and z''', and so on.

The nth order derivatives $y^{(n)}$ and $z^{(n)}$ will be defined here by a system of the form:

$$\begin{aligned} &\varphi'_y(x, y, z) \cdot y^{(n)} + \varphi'_z(x, y, z) \cdot z^{(n)} + A = 0 \,, \\ &\psi'_y(x, y, z) \cdot y^{(n)} + \psi'_z(x, y, z) \cdot z^{(n)} + B = 0 \,, \end{aligned} \qquad (17_1)$$

where A and B are expressions containing derivatives of orders lower than n. By elementary algebra, such a system will give a unique solution if the condition is satisfied:

$$\varphi'_y(x, y, z) \cdot \psi'_z(x, y, z) - \varphi'_z(x, y, z) \cdot \psi'_y(x, y, z) \neq 0 .$$

The method described above leads to fully defined values of the derivatives for all x, y, z for which this condition is fulfilled, and which satisfy system (17_1).

If a system of m equations in $(m + n)$ variables is given, the system generally speaking defines m variables as implicit functions of the remaining n variables, and the derivatives of these implicit functions can be obtained by the above method as a result of differentiation of the equations with respect to the independent variables.

158. Example. We take as an example the equation

$$ax^2 + by^2 + cz^2 = 1 \,, \qquad (18)$$

which defines z as a function of x and y. We get on differentiating with respect to x:

$$ax + cz \cdot z'_x = 0 \,, \qquad (19)$$

and similarly, differentiating with respect to y:

$$by + cz \cdot z'_y = 0 \,, \qquad (19_1)$$

whence

$$z'_x = -\frac{ax}{cz} \,, \quad z'_y = -\frac{by}{cz} \,.$$

We obtain on differentiating (19) with respect to x, and with respect to y, and (19_1) with respect to y:

$$a + cz'^2_x + cz \cdot z''_{x^2} = 0, \;\; cz'_x z'_y + cz \cdot z''_{yx} = 0, \;\; b + cz'^2_y + cz \cdot z''_{y^2} = 0$$

whence:

$$z''_{x^2} = -\frac{a + cz'^2_x}{cz} = -\frac{a + c\dfrac{a^2x^2}{c^2z^2}}{cz} = -\frac{acz^2 + a^2x^2}{c^2z^3},$$

$$z''_{xy} = -\frac{z'_x z'_y}{z} = -\frac{abxy}{c^2z^3},$$

$$z''_{y^2} = -\frac{b + cz'^2_y}{cz} = -\frac{bcz^2 + b^2y^2}{c^2z^3}.$$

We now give a second method of finding the partial derivatives, using the expression for the total differential of a function. We first prove an auxiliary theorem. Suppose we are asked to find an expression for the total differential dz of a function of two independent variables x and y in the form:

$$\mathrm{d}z = p\mathrm{d}x + q\mathrm{d}y.$$

We already know that

$$\mathrm{d}z = z'_x \,\mathrm{d}x + z'_y \,\mathrm{d}y.$$

We find, on comparing these two equations:

$$p\mathrm{d}x + q\mathrm{d}y = z'_x\mathrm{d}x + z'_y \,\mathrm{d}y.$$

But dx and dy are constants, being the differentials of the independent variables. We put d$x = 1$ and d$y = 0$, or d$x = 0$ and d$y = 1$, and find

$$p = z'_x \quad \text{and} \quad q = z'_y.$$

Thus, *if the total differential of a function z of two independent variables x and y can be put in the form :*

$$\mathrm{d}z = p\mathrm{d}x + q\mathrm{d}y,$$

we have $p = z'_x$ and $q = z'_y$.

This theorem is valid for a function of any number of independent variables. It can be shown similarly that, *if the second order differential can be put in the form :*

$$\mathrm{d}^2z = r\mathrm{d}x^2 + 2s\mathrm{d}x\mathrm{d}y + t\mathrm{d}y^2,$$

we have $r = z''_{x^2}$, $s = z''_{xy}$ and $t = z''_{y^2}$.

We now return to our example. Instead of finding the derivatives of the left-hand side of (18) with respect to x and y, we find its differential, whilst recalling that the expression for the first differential is independent of the choice of independent variables [153]:

$$ax\mathrm{d}x + by\mathrm{d}y + cz\mathrm{d}z = 0, \tag{20}$$

whence

$$\mathrm{d}z = -\frac{ax}{cz}\,\mathrm{d}x - \frac{by}{cz}\,\mathrm{d}y,$$

and therefore, by the theorem just proved:

$$z'_x = -\frac{ax}{cz} \quad \text{and} \quad z'_y = -\frac{by}{cz}.$$

We now find the differential of the left-hand side of (20), remembering that dx and dy have to be regarded as constants here:

$$a dx^2 + b dy^2 + c dz^2 + cz d^2 z = 0$$

or

$$d^2z = -\frac{a}{cz} dx^2 - \frac{b}{cz} dy^2 - \frac{1}{z} dz^2 = -\frac{a}{cy} dx^2 - \frac{b}{cz} dy^2 -$$

$$-\frac{1}{z}\left(\frac{ax}{cz} dx + \frac{by}{cz} dy\right)^2 = -\frac{acz^2 + a^2x^2}{a^2z^3} dx^2 - 2\frac{abxy}{c^2z^3} dx dy - \frac{bcz^2 + b^2y^2}{c^2z^3} dy^2,$$

and therefore:

$$z''_{x^2} = -\frac{acz^2 + a^2x^2}{c^2z^3}, \quad z''_{xy} = -\frac{abxy}{c^2z^3}, \quad z''_{y^2} = -\frac{bcz^2 + b^2y^2}{c^2z^3}.$$

Thus, having found the differential of a certain order, we can obtain all the partial derivatives of the corresponding order.

159. Existence of implicit functions. Our discussions have been of a formal kind, since we have presupposed in every case that the equation or system of equations implicitly defines a certain function that possesses derivatives. We now prove the basic existence theorem for implicit functions.

We take the equation

$$F(x, y) = 0 \tag{21}$$

and indicate under what conditions it uniquely defines y as a continuous function of x that possesses derivatives.

THEOREM. *Let $x = x_0$ and $y = y_0$ be solutions of equation* (21), i.e.

$$F(x_0, y_0) = 0; \tag{22}$$

let $F(x, y)$ and its partial derivatives of the first order with respect to x and y be continuous for all x, y sufficiently close to x_0, y_0, and finally, let the partial derivative $F'_y(x, y)$ differ from zero for $x = x_0$, $y = y_0$. In these circumstances, there exists a uniquely defined function $y(x)$ for all x sufficiently close to x_0, which satisfies equation (21), *is continuous, possesses a derivative and satisfies the condition: $y(x_0) = y_0$.*

We take for clarity, $F'(x, y) > 0$ at $x = x_0$, $y = y_0$. Since this derivative is continuous by hypothesis, it will also be positive for all x, y sufficiently near x_0, y_0, i.e. there exists a positive number l such that $F(x, y)$ and its partial derivatives are continuous, and

$$F'_y(x, y) > 0, \tag{23}$$

for all x, y, satisfying the condition:

$$|x - x_0| \leqslant l, \ |y - y_0| \leqslant l. \tag{24}$$

Moreover, by (22), function $F(x_0, y)$ of the single variable y vanishes for $y = y_0$, whilst by (23) and (24), it is an increasing function of y in the interval $(y_0 - l, y_0 + l)$. The numbers $F(x_0, y_0 - l)$ and $F(x_0, y_0 + l)$ thus have different signs: the former is negative, the latter positive. We can assert, on taking into account the continuity of $F(x, y)$ [67], that $F(x, y_0 - l)$ will be negative, and $F(x, y_0 + l)$ positive, for all x sufficiently near x_0, i.e. there exists a positive number l_1, such that

$$F(x, y_0 - l) < 0 \text{ and } F(x, y_0 + l) > 0 \tag{25}$$

for $|x - x_0| \leqslant l_1$. Let m denote the smaller of l, l_1. We can say, from (24) and (25), that inequalities (23) and (25) are satisfied if x and y satisfy the inequalities:

$$|x - x_0| \leqslant m, \ |y - y_0| \leqslant l. \tag{26}$$

If we take any definite x, lying in the interval $(x_0 - m, x_0 + m)$, i.e. satisfying the first of inequalities (26), we can see by (23) that $F(x, y)$, considered as a function of y, will be increasing in the interval $(y_0 - l, y_0 + l)$, and by (25), it will have different signs at the ends of this interval. It will therefore vanish for a uniquely defined value of y in the interval. In particular, if $x = x_0$, this value of y will be $y = y_0$, by (22). We have thus proved the existence of a definite function $y(x)$ in the interval $(x_0 - m, x_0 + m)$, representing a solution of equation (21) and satisfying the condition $y(x_0) = y_0$. In other words, it follows from the foregoing arguments that equation (21) has a unique root, inside the interval $(y_0 - l, y_0 + l)$, for every fixed x from the interval $(x_0 - m, x_0 + m)$.

We now show that the function obtained, $y(x)$, is continuous at $x = x_0$. In fact, given any small positive ε, the numbers $F(x_0, y_0 - \varepsilon)$ and $F(x_0, y_0 + \varepsilon)$ will have different signs, by (25): so that a positive η will exist, such that $F(x, y_0 - \varepsilon)$ and $F(x, y_0 + \varepsilon)$ have different signs, provided only that $|x - x_0| < \eta$. Thus, given $|x - x_0| < \eta$, the root of equation (21), i.e. the value of our function $y(x)$, satisfies $|y - y_0| < \varepsilon$, which proves the continuity of $y(x)$ at $x = x_0$.

We now prove the existence of the derivative $y'(x)$ at $x = x_0$. Let $x - x_0 = \Delta x$, and let the corresponding increment of y be $\Delta y = y - y_0$. Hence, $x = x_0 + \Delta x$ and $y = y_0 + \Delta y$ satisfy (21), i.e. we have $F(x_0 + \Delta x, y_0 + \Delta y) = 0$, and we can write, by (22):

$$F(x_0 + \Delta x, y_0 + \Delta y) - F(x_0, y_0) = 0.$$

We can use the continuity of the partial derivatives to rewrite this equation as [68]:

$$[F'_{x_0}(x_0, y_0) + \varepsilon_1]\,\Delta x + [F'_{y_0}(x_0, y_0) + \varepsilon_2]\,\Delta y = 0\,, \tag{27}$$

where ε_1 and $\varepsilon_2 \to 0$ if Δx and $\Delta y \to 0$, and where $F'_{x_0}(x_0, y_0)$ and $F'_{y_0}(x_0, y_0)$ denote the values of the partial derivatives at $x = x_0$, $y = y_0$. It follows from the continuity proved above that $\Delta y \to 0$ if $\Delta x \to 0$.

Equation (27) gives us:

$$\frac{\Delta y}{\Delta x} = -\frac{F'_{x_0}(x_0, y_0) + \varepsilon_1}{F'_{y_0}(x_0, y_0) + \varepsilon_2}\,;$$

we pass to the limit as $\Delta x \to 0$, and get:

$$y'(x_0) = -\frac{F'_{x_0}(x_0, y_0)}{F'_{y_0}(x_0, y_0)}.$$

We have proved the continuity of $y(x)$, and the existence of its derivative, only for $x = x_0$. On taking any other value of x from the interval $(x_0 - m, x_0 + m)$, and the corresponding value of y from the interval $(y_0 - l, y_0 + l)$, representing the root of (21), all the conditions of our theorem are again satisfied for this pair of values of x and y, and by what has been proved, $y(x)$ will be continuous and will have a derivative for the value of x concerned.

We can state and prove, in exactly the same way, a theorem regarding the existence of the implicit function $z(x, y)$, defined by the equation:

$$\Phi(x, y, z) = 0.$$

We now consider the system:

$$\varphi(x, y, z) = 0, \quad \psi(x, y, z) = 0, \tag{28}$$

defining y and z as functions of x.

The following theorem holds in this case:

THEOREM. *Let $x = x_0$, $y = y_0$, $z = z_0$ be solutions of system* (28); *let $\varphi(x, y, z)$, $\psi(x, y, z)$ and their first order partial derivatives be continuous functions of (x, y, z) for all values of these variables sufficiently near (x_0, y_0, z_0); and let the expression*

$$\varphi'_y(x, y, z)\,\psi'_z(x, y, z) - \varphi'_z(x, y, z)\,\psi'_y(x, y, z)$$

differ from zero for $x = x_0, y = y_0, z = z_0$. In these circumstances, there exists for all x sufficiently near x_0 a uniquely defined system of two functions $y(x)$, $z(x)$, that satisfies equations (28); *these functions are continuous, have first order derivatives, and satisfy the conditions: $y(x_0) = y_0$, $z(x_0) = z_0$.*

We shall not dwell on the proof of this theorem. The general case of any number of functions with any number of variables is considered in the third volume.

160. Curves in space and surfaces. We know from analytic geometry that, generally speaking, for every equation with three variables

$$F(x, y, z) = 0, \tag{29}$$

or explicitly,

$$z = f(x, y), \tag{30}$$

there is a corresponding surface in space, relative to rectangular axes OX, OY, OZ.

A line in space can be considered as the intersection of two surfaces, and can therefore be defined by a combination of two equations:

$$F_1(x, y, z) = 0, \quad F_2(x, y, z) = 0. \tag{31}$$

Alternatively, a curve can be defined by parametric equations:

$$x = \varphi(t),\ y = \psi(t),\ z = \omega(t). \tag{32}$$

The length of arc of a curve is defined, as in the case of a plane curve, as the limit of the perimeter of a series of joined chords inscribed in the arc, on indefinite decrease of the length of each chord. We omit the proof, since it follows exactly the same lines as in [103] for the case of a plane curve, that the length of arc is expressed by the definite integral:

$$s = \int_{(M_1)}^{(M_2)} \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2} = \int_{t_1}^{t_2} \sqrt{\varphi'^2(t) + \psi'^2(t) + \omega'^2(t)}\, \mathrm{d}t, \tag{33}$$

where t_1 and t_2 are the values of the parameter t corresponding to the ends M_1 and M_2 of the arc, and the differential of the arc has the expression:

$$\mathrm{d}s = \sqrt{(\mathrm{d}x)^2 + (\mathrm{d}y)^2 + (\mathrm{d}z)^2}. \tag{34}$$

If the role of parameter t is played by the length of arc s, measured from some given point of the arc, it can be shown, by an exactly similar method to that used in the case of a plane curve [70], that the derivatives $\mathrm{d}x/\mathrm{d}s$, $\mathrm{d}y/\mathrm{d}s$, $\mathrm{d}z/\mathrm{d}s$ are equal to the direction cosines of the tangent to the curve, i.e. are equal to the cosines of the angles formed by the positive direction of the tangent with the coordinate axes. Thus, *the direction cosines of the tangent to the curve at the point* (x, y, z) *of the curve,* i.e. *the cosines of the angles formed by the direction of the tangent with the coordinate axes, are proportional to* $\mathrm{d}x$, $\mathrm{d}y$ *and* $\mathrm{d}z$, and the equation of the tangent can be written in the form:

$$\frac{X - x}{\mathrm{d}x} = \frac{Y - y}{\mathrm{d}y} = \frac{Z - z}{\mathrm{d}z}, \tag{35}$$

or

$$\frac{X - \varphi(t)}{\varphi'(t)} = \frac{Y - \psi(t)}{\psi'(t)} = \frac{Z - \omega(t)}{\omega'(t)}. \tag{36}$$

We now introduce a new concept, that of *the tangent plane to the surface:*

$$F(x, y, z) = 0. \tag{37}$$

Let $M(x, y, z)$ be a point of the surface, and let L be a line on the surface passing through the point M. The coordinates of points

of the line are functions of some parameter t, whilst the functions satisfy equation (37), since the line L lies on the surface. Equation (37) must be satisfied along the whole of line L, i.e. for all values of t, and we can write here, on differentiating the left-hand side of (37):

$$F'_x(x, y, z)\,dx + F'_y(x, y, z)\,dy + F'_z(x, y, z)\,dz = 0. \tag{38}$$

We assume that $F(x, y, z)$ has continuous partial derivatives F'_x, F'_y, F'_z, at least one of which is non-zero.

But we know from analytic geometry that an equation of the form

$$aa_1 + bb_1 + cc_1 = 0$$

is the condition for two straight lines, one with direction cosines proportional to a, b, c, and the other with direction cosines proportional to a_1, b_1, c_1, to be perpendicular. As we have seen, dx, dy, dz are proportional to the direction-cosines of the tangent to the line L at the point M. Equation (38) thus shows that the tangent to the line L at the point M is perpendicular to some specific direction, independent of L, the direction cosines of which are proportional to $F'_x(x,y,z)$, $F'_y(x,y,z)$, and $F'_z(x, y, z)$. We see that *the tangents to every line L, lying on the surface* (37), *and passing through the point M, lie in the same plane*

$$A(X - x) + B(Y - y) + C(Z - z) = 0, \tag{39}$$

which is referred to as the tangent plane to the surface at the point M.

We know from analytic geometry that the coefficients A, B, C in the equation of the plane are proportional to the direction cosines of the normal to the plane, i.e. proportional to $F'_x(x, y, z)$, $F'_y(x, y, z)$, $F'_z(x, y, z)$ in the present case; so that the equation of the tangent plane can finally be written as:

$$F'_x(x, y, z)\,(X - x) + F'_y(x, y, z)\,(Y - y) + F'_z(x, y, z)\,(Z - z) = 0, \tag{40}$$

where X, Y, Z are the current coordinates of the tangent plane and x, y, z are the coordinates of the point of contact M.

The normal to the tangent plane, passing through the point of contact M, is called the normal to the surface. Its direction cosines are proportional to the partial derivatives $F'_x(x, y, z)$, $F'_y(x, y, z)$, $F'_z(x, y, z)$, as we have just seen, and its equation is therefore:

$$\frac{X - x}{F'_x(x, y, z)} = \frac{Y - y}{F'_y(x, y, z)} = \frac{Z - z}{F'_z(x, y, z)}. \tag{41}$$

If the surface is given by the explicit equation: $z = f(x, y)$, equation (37) takes the form:

$$F(x, y, z) = f(x, y) - z = 0\,,$$

and therefore:

$$F'_x(x, y, z) = f'_x(x, y),\ \ F'_y(x, y, z) = f'_y(x, y),\ \ F'_z(x, y, z) = -1\,.$$

If we adopt the usual notation of p for $f'_x(x, y)$, and q for $f'_y(x, y)$, we obtain for the equation of the tangent plane:

$$p(X - x) + q(Y - y) - (Z - z) = 0 \tag{42}$$

and for the normal to the surface:

$$\frac{X - x}{p} = \frac{Y - y}{q} = \frac{Z - z}{-1}\,. \tag{43}$$

In the case of the ellipsoid

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2} = 1$$

the equation of the tangent plane at any given point of (x, y, z) is:

$$\frac{2x}{a^2}(X - x) + \frac{2y}{b^2}(Y - y) + \frac{2z}{c^2}(Z - z) = 0$$

or

$$\frac{xX}{a^2} + \frac{yY}{b^2} + \frac{zZ}{c^2} = \frac{x^2}{a^2} + \frac{y^2}{b^2} + \frac{z^2}{c^2}\,.$$

The right-hand side of this equation is simply equal to unity, since the coordinates (x, y, z) of the point of contact must satisfy the equation of the ellipsoid; thus the equation of the tangent plane is finally:

$$\frac{xX}{a^2} + \frac{yY}{b^2} + \frac{zZ}{c^2} = 1\,.$$

§ 16. Taylor's formula. Maxima and minima of functions of several variables

161. Extension of Taylor's formula to functions of several independent variables. We confine ourselves to the function $f(x, y)$ of two independent variables, so as to simplify the writing. *Taylor's formula* gives the expansion of $f(a + h, b + k)$ in powers of the increments h and k

of the independent variables [127]. We introduce a new independent variable t by putting:

$$x = a + ht, \; y = b + kt. \tag{1}$$

This gives us a function of the single independent variable t:

$$\varphi(t) = f(x, y) = f(a + ht, b + kt) \,,$$

where

$$\varphi(0) = f(a, b) \text{ and } \varphi(1) = f(a + h, b + k). \tag{2}$$

We can write, using Maclaurin's formula with Lagrange's remainder term [127]:

$$\varphi(1) = \varphi(0) + \frac{\varphi'(0)}{1!} + \frac{\varphi''(0)}{2!} + \ldots +$$

$$+ \frac{\varphi^{(n)}(0)}{n!} + \frac{\varphi^{(n+1)}(\theta)}{(n+1)!} \quad (0 < \theta < 1). \tag{3}$$

We now express the derivatives $\varphi^{(p)}(0)$ and $\varphi^{(n+1)}(\theta)$ in terms of the function $f(x, y)$.

From (1), x and y are linear functions of the independent variable t, and

$$\mathrm{d}x = h\mathrm{d}t, \; \mathrm{d}y = k\mathrm{d}t.$$

We can thus use our symbolic formula for finding the differential of any order of function $\varphi(t)$ [156]:

$$\mathrm{d}^p \varphi(t) = \left(\frac{\partial}{\partial x}\mathrm{d}x + \frac{\partial}{\partial y}\mathrm{d}y\right)^{(p)} f(x, y) = \left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^{(p)} f(x, y)\,\mathrm{d}t^p,$$

whence

$$\varphi^{(p)}(t) = \frac{\mathrm{d}^p \varphi(t)}{\mathrm{d}t^p}\left(h\frac{\partial}{\partial x} + k\frac{\partial}{\partial y}\right)^{(p)} f(x, y).$$

We have, for $t = 0$, $x = a$ and $y = b$, and for $t = \theta$, $x = a + \theta h$ and $y = b + \theta k$, and hence:

$$\varphi^{(p)}(0) = \left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b}\right)^{(p)} f(a, b)\,,$$

$$\varphi^{(n+1)}(\theta) = \left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b}\right)^{(n+1)} f(a + \theta h, \; b + \theta k).$$

We substitute these expressions in (3), and make use of (2) again, and finally obtain Taylor's formula:

$$f(a+h, b+k) = f(a, b) + \left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b}\right) f(a, b) +$$
$$+ \frac{1}{2!}\left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b}\right)^{(2)} f(a, b) + \ldots + \frac{1}{n!}\left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b}\right)^{(n)} f(a, b) +$$
$$+ \frac{1}{(n+1)!}\left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b}\right)^{(n+1)} f(a+\theta h,\ b+\theta k). \tag{4}$$

If we replace a and b by x and y respectively in this formula, denote the increments h and k of the independent variables by $\mathrm{d}x$ and $\mathrm{d}y$, and denote the increment of the function, i.e. $f(x + \mathrm{d}x, y + \mathrm{d}y) - f(x, y)$, by $\Delta f(x, y)$, we can write the formula as follows:

$$\Delta f(x, y) = \mathrm{d}f(x, y) + \frac{\mathrm{d}^2 f(x, y)}{2!} + \ldots + \frac{\mathrm{d}^n f(x, y)}{n!} + \left[\frac{\mathrm{d}^{n+1} f(x, y)}{(n+1)!}\right]_{\substack{x+\theta\,\mathrm{d}x \\ y+\theta\,\mathrm{d}y}}.$$

The right-hand side of this formula contains the differentials of various orders of $f(x, y)$, whilst the values that have to be substituted for the independent variables in the $(n + 1)$th order derivative are indicated for the last term. As in the case of a function of a single independent variable, we can deduce Maclaurin's formula, giving an expansion of $f(x, y)$ in powers of x and y, by setting in Taylor's formula (4):

$$a = 0,\ b = 0;\ h = x,\ k = y.$$

We assumed, when obtaining (4), that $f(x, y)$ has continuous partial derivatives up to order $(n + 1)$ in some open domain that contains the straight line joining points (a, b) and $(a + h,\ b + k)$. The point $x = a + ht$, $y = b + kt$, describes this line as t varies from zero to unity. We obtain the formula for finite increments, with $n = 0$:

$$f(a + h, b + k) - f(a, b) = hf'_a(a + \theta\, h, b + \theta\, k) +$$
$$+ kf'_b(a + \theta\, h, b + \theta\, k).$$

It immediately follows from this, as in [63], that *if the first order partial derivatives are everywhere zero in a certain domain, the function maintains a constant value in the domain.*

162. Necessary conditions for maxima and minima of functions. Let function $f(x, y)$ be continuous at, and in some neighbourhood of, the point (a, b). We shall say, by analogy with the case of a single

independent variable, that *the function $f(x, y)$ of two independent variables attains a maximum at the point (a, b), if $f(a, b)$ is not less than all adjacent values of the function*, i.e. *if*

$$\Delta f = f(a + h, b + k) - f(a, b) \leqslant 0 \,, \tag{5}$$

for all h and k that are sufficiently small in absolute value.

Similarly, we say that $f(x, y)$ *attains a minimum at $x = a$, $y = b$*, if

$$\Delta f = f(a + h, b + k) - f(a, b) \geqslant 0 \tag{5_1}$$

for all h and k sufficiently small in absolute value.

Now, let $x = a$, $y = b$ be values of the independent variables for which function $f(x, y)$ attains a maximum or minimum. We consider the function of a single independent variable $f(x, b)$. By hypothesis, it must attain a maximum or minimum for $x = a$, so that its derivative with respect to x must either vanish or not exist at $x = a$ [58]. Using the same argument, we see that the derivative of $f(a, y)$ with respect to y must either vanish or not exist at $y = b$. This leads us to the following necessary condition for a maximum or minimum: *the function of two independent variables $f(x, y)$ can only attain a maximum or minimum at values of x and y for which the first order derivatives $\partial f(x, y)/\partial x$ and $\partial f(x, y)/\partial y$ either vanish or do not exist.*

Similarly, letting either only x or only y vary, we can assert by what was said in [58] that, given the existence of the second order derivatives, a necessary condition for a maximum is given by $\partial^2 f(x,y)/\partial x^2 \leqslant 0$ and $\partial^2 f(x, y)/\partial y^2 \leqslant 0$, and a necessary condition for a minimum by $\partial^2 f(x, y)/\partial x^2 \geqslant 0$ and $\partial^2 f(x, y)/\partial y^2 \geqslant 0$.

The above reasoning remains applicable for functions of any number of independent variables. We can thus state the following general rule:

A function of several independent variables can attain a maximum or minimum only for those values of the independent variables for which the first order derivatives either vanish or do not exist. We shall in future limit ourselves to considering cases when the partial derivatives mentioned exist.

The first order differential is equal to the sum of the products of the partial derivatives with respect to the independent variables and the differentials of the corresponding independent variables [153], so that we can say that *the first order differential of a function must vanish at the values of the independent variables for which the function has a maximum or minimum.* This form of necessary condition is con-

venient, since the expression for the first differential does not depend on the choice of variables [153]. By equating the first order partial derivatives to zero, we get a system of equations, defining the values of the independent variables for which the function can attain maxima or minima. A complete solution requires further investigation to find out if the function in fact attains maxima or minima at the values obtained of the independent variables, and if it does, to say which of the two it attains, a maximum or a minimum. We indicate in the next section how the investigation proceeds for a function of two independent variables.

163. Investigation of the maxima and minima of a function of two independent variables. Let the system of equations

$$\frac{\partial f(x, y)}{\partial x} = 0, \quad \frac{\partial f(x, y)}{\partial y} = 0, \tag{6}$$

expressing the necessary condition for a maximum or minimum, give us the values $x = a$, $y = b$, that are to be investigated. We assume that $f(x, y)$ has continuous partial derivatives of the second order at and in some neighbourhood of the point (a, b).

We can write, in accordance with Taylor's formula (4), with $n = 2$:

$$f(a+h, b+k) = f(a, b) + \frac{\partial f(a, b)}{\partial a} h + \frac{\partial f(a, b)}{\partial b} k +$$
$$+ \frac{1}{2!}\left[\frac{\partial^2 f(x, y)}{\partial x^2} h^2 + 2\frac{\partial^2 f(x, y)}{\partial x \partial y} hk + \frac{\partial^2 f(x, y)}{\partial y^2} k^2\right]_{\substack{x=a+\theta h \\ y=b+\theta k}}.$$

We use the fact that $x = a$, $y = b$ represents a solution of system (6), in order to rewrite the above equation as:

$$\Delta f = f(a+h, b+k) - f(a, b) =$$
$$= \frac{1}{2!}\left[\frac{\partial^2 f(x, y)}{\partial x^2} h^2 + 2\frac{\partial^2 f(x, y)}{\partial x \partial y} hk + \frac{\partial^2 f(x, y)}{\partial y^2} k^2\right]_{\substack{x=a+\theta h \\ y=b+\theta k}}. \tag{7}$$

We put:

$$r = \sqrt{h^2 + k^2}, \quad h = r\cos\alpha, \quad k = r\sin\alpha.$$

When h and k are small in absolute value, r is also small, and vice versa; the condition that h and $k \to 0$, and the condition that $r \to 0$, are thus equivalent.

We now have for (7):

$$\Delta f = \frac{r^2}{2!}\left[\frac{\partial^2 f(x, y)}{\partial x^2}\cos^2 \alpha + 2\,\frac{\partial^2 f(x, y)}{\partial x\,\partial y}\cos\alpha\,\sin\alpha + \right.$$
$$\left. + \frac{\partial^2 f(x, y)}{\partial y^2}\sin^2\alpha\right]_{\substack{x=a+\theta h\\ y=b+\theta k}}. \tag{8}$$

We can say by the continuity of the second order derivatives, on taking h and k, or what comes to the same thing, r, as infinitesimals, that evaluation of the derivatives on the right-hand side of (8) for $a + \theta h$, $b + \theta k$, which differ infinitesimally from a, b, gives results that themselves differ infinitesimally from the numbers:

$$\frac{\partial^2 f(a, b)}{\partial a^2} = A\,, \quad \frac{\partial^2 f(a, b)}{\partial a\,\partial b} = B\,, \quad \frac{\partial^2 f(a\,b)}{\partial b^2} = C\,;$$

the coefficients of $\cos^2\alpha$, $\cos\alpha\sin\alpha$, and $\sin^2\alpha$ in the square brackets in (8) can therefore be replaced by:

$$A + \varepsilon_1,\ 2B + \varepsilon_2,\ C + \varepsilon_3$$

respectively, where ε_1, ε_2, and ε_3 are infinitesimals along with h and k (or r).

We can now write (8) in the form:

$$\Delta f = \frac{r^2}{2!}\left[A\cos^2\alpha + 2B\sin\alpha\cos\alpha + C\sin^2\alpha + \varepsilon\right]. \tag{9}$$

where

$$\varepsilon = \varepsilon_1\cos^2\alpha + 2\varepsilon_2\cos\alpha\sin\alpha + \varepsilon_3\sin^2\alpha$$

is an infinitesimal along with h and k (or r).

It follows from the definition of a maximum or minimum that, if the right-hand side of (9) preserves a $(-)$ sign for all sufficiently small r, the values $x = a$ and $y = b$ give a maximum of $f(x, y)$; whilst they give a minimum if the $(+)$ sign is preserved. Finally, if the right-hand side of (9) can have both a $(+)$ and a $(-)$ sign for arbitrarily small values of r, $x = a$ and $y = b$ give neither a maximum nor a minimum of the function.

There are four possible cases when investigating the sign of the right-hand side of (9):

I. If the expression

$$A\cos^2\alpha + 2B\sin\alpha\cos\alpha + C\sin^2\alpha \tag{10}$$

does not vanish for any value of α, it preserves the same sign, since it is a continuous function of α [55]. Let the sign preserved be $(+)$.

It must attain a least (positive) value in the interval $(0, 2\pi)$, say m, which will hold for any a, due to the periodicity of $\sin a$ and $\cos a$. We shall have $|\varepsilon|$ less than m for all sufficiently small r, and the sign of the right-hand side of (9) in this case will be given by the sign of (10), i.e. will be $(+)$. We thus have a minimum in this case.

II. Now let (10) preserve the $(-)$ sign, whilst not vanishing for any a. Let m be the greatest (negative) value of (10), for a varying in the interval $(0, 2\pi)$. We have $|\varepsilon|$ less than m for sufficiently small r, in which case the sign of the right-hand side of (9) remains $(-)$, i.e. we have a maximum here.

III. We now suppose that (10) changes sign. Let it be equal to the positive number $+m_1$ for $a = a_1$, and equal to the negative number $-m_2$ for $a = a_2$. We shall have $|\varepsilon|$ less than m_1 and m_2 for all sufficiently small r, and with these r, and $a = a_1$ or a_2, the sign of the right-hand side of (9) is determined by the sign of (10), i.e. is $(+)$ for $a = a_1$ and $(-)$ for $a = a_2$. It follows that the sign of the right of (9) can here be both $(+)$ and $(-)$ for arbitrarily small r, i.e. we have neither a maximum nor a minimum in this case.

IV. We finally suppose that (10), whilst preserving an invariable sign, can vanish for several values of a. Further investigation would be needed into the sign of ε, for us to be able to draw any conclusions regarding the sign of the right-hand side of (9); so that this case remains doubtful in our present study.

All the above leads to investigation of the sign of (10) as a varies, and we indicate some simple tests that enable us to decide with which of the four cases we are concerned.

1. We take $A \neq 0$ to start with. We can write (10) in the form:

$$\frac{(A\cos a + B\sin a)^2 + (AC - B^2)\sin^2 a}{A}. \tag{11}$$

If $AC - B^2 > 0$, the numerator of fraction (11) consists of the sum of two positive terms, which cannot vanish simultaneously. The second term vanishes, in fact, only when $\sin a = 0$, in which case $\cos a = \pm 1$, and the first term becomes $A^2 \neq 0$. The sign of (11) is thus the same as the sign of A in the present case, so that we have case (I) with $A > 0$, i.e. a minimum, and case (II) with $A < 0$, i.e. a maximum.

2. We take $A \neq 0$ as before, and suppose that $AC - B^2 < 0$. The numerator of (11) will have the $(+)$ sign with $\sin a = 0$, and the $(-)$ sign with $\cot a = -B/A$, so that we obtain case (III) in these circumstances, i.e. neither a maximum nor a minimum.

3. If we take $AC - B^2 = 0$, with $A \neq 0$, the numerator of (11) reduces to the first term; it retains the invariable (+) sign, whilst vanishing at $\cot \alpha = -(A/B)$, i.e. we have here the doubtful case (IV).

4. We take $A = 0$, $B \neq 0$. The expression (10) now has the form: $\sin \alpha\ (2B \cos \alpha + C \sin \alpha)$. The expression in brackets preserves an invariable sign for α close to zero, which is the same as the sign of B; whereas the first factor, $\sin \alpha$, has different signs depending on whether α is greater or less than zero. We thus have case (III) here: neither a maximum nor a minimum.

5. We finally take $A = B = 0$. Expression (10) now reduces to the single term $C \sin^2 \alpha$. It can therefore vanish, whilst not changing sign, i.e. we have the doubtful case.

Noticing that $AC - B^2 < 0$ in case 4 above, whilst $AC - B^2 = 0$ in case 5, we are able to state the following rule:

The procedure for finding the maxima and minima of a function $f(x, y)$ of two independent variables x and y is to obtain the partial derivatives $f'_x(x, y)$ and $f'_y(x, y)$ and solve the system of equations :

$$f'_x(x, y,) = 0,\ f'_y(x, y) = 0.$$

Let $x = a$, $y = b$ be a solution of this system. We put

$$\frac{\partial^2 f(a, b)}{\partial a^2} = A, \quad \frac{\partial^2 f(a, b)}{\partial a\, \partial b} = B, \quad \frac{\partial^2 f(a, b)}{\partial b^2} = C,$$

and investigate the solution in accordance with the following arrangement :

$AC-B^2$	+		−	0
A	+	−	no min. no max.	doubtful case
	min.	max.		

164. Examples. 1. We consider the surface $z = f(x, y)$. The equation of the tangent plane to it is [160]:

$$p(X - x) + q(Y - y) - (Z - z) = 0,$$

where p and q denote the partial derivatives $f'_x(x, y)$, $f'_y(x, y)$.

If z attains a maximum or minimum for certain values $x = a$, $y = b$, the corresponding point is called a *vertex* of the surface; the tangent plane at such a point must be parallel to the XY plane, i.e. the partial derivatives p and q must vanish, and the surface must be situated on one side of the

tangent plane near the point of contact (Fig. 163). It may happen, however, that p and q vanish at a certain point, i.e. the tangent plane is parallel to the XY plane, yet the the surface is situated on both sides of the tangent plane in the neighbourhood of the point; in this case, z attains neither a maximum nor a minimum for the x, y concerned.

We mention the further possibility, that comes under our heading of a "doubt ful case". We suppose that the tangent plane is parallel to the XY plane at $x = a$, $y = b$, and that the surface, whilst lying on one side of the tangent plane, has a line in common with it, passing through the point of contact. Here, although the difference

$$f(a + h, b + k) - f(a, b)$$

does not change sign for h and k that are sufficiently small in absolute value, it vanishes for non-zero h or k. This case is easily obtained, as, for instance, with a circular cylinder, the axis of which is parallel to the XY plane. The function $f(x, y)$ is also spoken of here as having a maximum or a minimum at $x = a$, $y = b$.

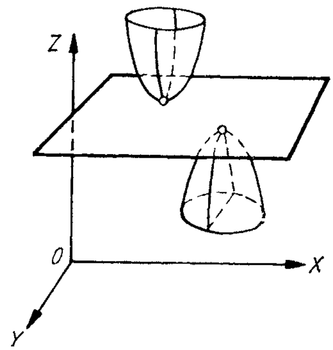

FIG. 163

The surface

$$2z = \frac{x^2}{a^2} - \frac{y^2}{b^2}$$

is a hyperbolic paraboloid. Equating to zero the partial derivatives of z with respect to x and y gives us $x = y = 0$, and the tangent-plane to the surface at the origin coincides with the XY plane. We obtain the second order partial derivatives:

$$\frac{\partial^2 z}{\partial x^2} = \frac{1}{a^2}, \quad \frac{\partial^2 z}{\partial x\, \partial y} = 0, \quad \frac{\partial^2 z}{\partial y^2} = -\frac{1}{b^2},$$

and therefore,

$$AC - B^2 = -\frac{1}{a^2 b^2} < 0,$$

i.e. z attains neither a maximum nor a minimum at $x = y = 0$, and the surface is situated on both sides of the tangent plane near the origin (Fig. 164).

2. We are given n points $M_i(a_i, b_i)$ $(i = 1, 2, \ldots, n)$ on a plane. The problem is to find the point M, such that the sum of the products of certain given numbers m_i and the squares of the distances MM_i is a minimum.

Let (x, y) be the coordinates of the point M. The sum referred to is:

$$w = \sum_{i=1}^{n} m_i \left[(x - a_i)^2 + (y - b_i)^2\right].$$

Equating the partial derivatives w'_x and w'_y to zero gives:

$$x = \frac{m_1 a_1 + m_2 a_2 + \ldots + m_n a_n}{m_1 + m_2 + \ldots + m_n}; \quad y = \frac{m_1 b_1 + m_2 b_2 + \ldots + m_n b_n}{m_1 + m_2 + \ldots + m_n}. \quad (12)$$

It is easily shown that A and $AC - B^2$ are greater than zero here, so that the values found for x and y in fact give a minimum of w. This minimum is the least value of w in the (x, y) plane, since $w \to \infty$ as the point (x, y) moves to an infinite distance.

If M_i are material particles, of masses m_i, *formula* (12) *defines the coordinates of the centre of gravity of the system of particles* M_i.

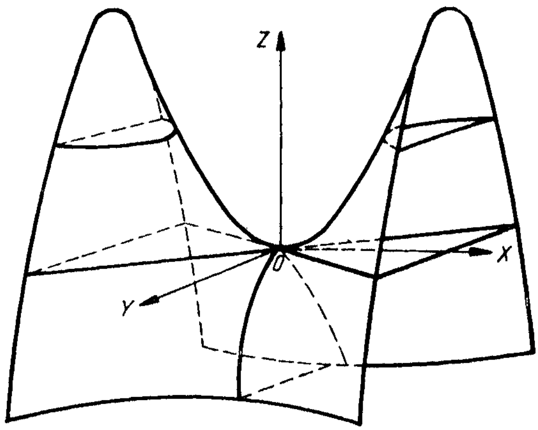

FIG. 164

165. Additional remarks on finding the maxima and minima of a function. The above discussion can be extended to the case of a larger number of independent variables. For instance, let a function $f(x, y, z)$ of three independent variables be given. The values of the independent variables, for which this function attains a maximum or a minimum, are found by solving the system of three equations with three unknowns [162]:

$$f_x'(x, y, z) = 0,\ f_y'(x, y, z) = 0,\ f_z'(x, y, z) = 0. \quad (13)$$

Let $x = a$, $y = b$, $z = c$ be one of the solutions of this system. We indicate briefly the way to investigate these values. Taylor's formula gives the increment of the function as a sum of homogeneous polynomials, arranged in powers of the increments of the independent variables:

$$\Delta f = h\frac{\partial f(a, b, c)}{\partial a} + k\frac{\partial f(a, b, c)}{\partial b} + l\frac{\partial f(a, b, c)}{\partial c} + \frac{1}{2!}\left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b} + l\frac{\partial}{\partial c}\right)^{(2)} f(a, b, c) + \ldots + \frac{1}{(n+1)!}\left(h\frac{\partial}{\partial a} + k\frac{\partial}{\partial b} + l\frac{\partial}{\partial c}\right)^{(n+1)} f(a + \theta h, b + \theta k, c + \theta l)$$

$$(0 < \theta < 1). \quad (14)$$

Since $x = a$, $y = b$, $z = c$ satisfy equations (13), we have

$$h \frac{\partial f(a, b, c)}{\partial a} + k \frac{\partial f(a, b, c)}{\partial b} + l \frac{\partial f(a, b, c)}{\partial c} = 0.$$

If the combination of terms to the second power in h, k, l,

$$\frac{1}{2!}\left(h \frac{\partial}{\partial a} + k \frac{\partial}{\partial b} + l \frac{\partial}{\partial c}\right)^{(2)} f(a, b, c) \tag{15}$$

does not vanish, the sign of the right-hand side of (14) is the same as that of expression (15) for h, k, l sufficiently small in absolute value, and if the sign is $(+)$, $f(a, b, c)$ is a minimum of $f(x, y, z)$, and if $(-)$, a maximum. If (15) can have different signs, $f(a, b, c)$ is neither a maximum nor a minimum. If, finally, (15) keeps its sign, but vanishes for some values of h, k, l, the case is a doubtful one, and further study is needed of the terms on the right-hand side of equation (14) that contain h, k, l, to higher powers than the second.

We carry out a full investigation of a doubtful case for a particular example of a function of two independent variables:

$$u = x^2 - 2xy + y^2 + x^3 + y^3.$$

The partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ vanish for $x = y = 0$. We have further:

$$A = \frac{\partial^2 u}{\partial x^2}\Bigg|_{\substack{x=0\\y=0}} = 2, \qquad B = \frac{\partial^2 u}{\partial x\, \partial y}\Bigg|_{\substack{x=0\\y=0}} = -2, \qquad C = \frac{\partial^2 u}{\partial y^2}\Bigg|_{\substack{x=0\\y=0}} = 2,$$

$$AC - B^2 = 0,$$

i.e. we are concerned with a doubtful case. It is a characteristic of this case that the combination of terms of the second degree in the expression for u consists of a perfect square, and we can write in this particular example:

$$u = (x - y)^2 + (x^3 + y^3).$$

When $x = y = 0$, u also vanishes. We introduce polar coordinates:

$$x = r \cos a, \quad y = r \sin a,$$

in order to investigate the sign of u for x and y near zero.

Substitution gives us:

$$u = r^2[(\cos a - \sin a)^2 + r(\cos^3 a + \sin^3 a)].$$

For any a, except $\pi/4$ and $5\pi/4$, in the interval $(0,2\pi)$,

$$\cos a - \sin a \neq 0,$$

so that we can choose, for such an a, a positive number r_0 such that the sign of the expression in square brackets is $(+)$ for $r < r_0$. The sign is likewise $(+)$ for $a = \pi/4$, but we get a $(-)$ sign for $a = 5\pi/4$, and u has therefore neither a maximum nor a minimum at $x = y = 0$.

We also consider the function

$$u = (y - x^2)^2 - x^5 .$$

It is easily shown that the partial derivatives $\partial u/\partial x$ and $\partial u/\partial y$ vanish at $x = y = 0$, and that we have a doubtful case. We find, on taking x arbitrarily small and putting $y = x^2$, that u becomes $(-x^5)$, so that its sign depends on the sign of x, i.e. u has neither a maximum nor a minimum for $x = y = 0$. Introducing polar coordinates would have given:

$$u = r^2(\sin^2 a - 2r \cos^2 a \sin a + r^2 \cos^4 a - r^3 \cos^5 a) ,$$

and it is evident from this expression that a positive number r_0 can be found for any a, not excluding $a = 0$ and π, such that $u > 0$ for $r < r_0$, i.e. u has a $(+)$ sign near the origin on any radius vector through the origin. But, as

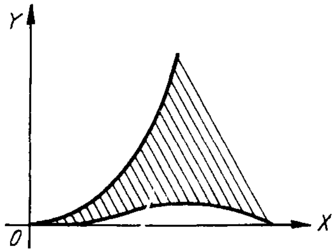

FIG. 165

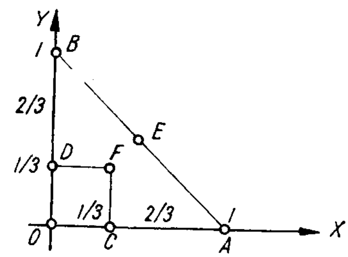

FIG. 166

we have seen, this does not in itself imply a minimum at the origin, where $u = 0$, since it is impossible to find an r_0 that is *the same* for all values of a.

We plotted the curve $(y - x^2)^2 - x^5 = 0$ in [76], and saw that it has a cusp of the second kind at the origin, whilst the left-hand side of this equation has a $(-)$ sign near the origin, if its value is considered at points lying inside the shaded region between the two branches of the curve (Fig. 165).

166. The greatest and least values of a function. Suppose that we wish to find the greatest value (absolute maximum) of some function $f(x, y)$, given in a certain domain. The procedure of [163] enables us to find all the maxima of the function *within* this domain, i.e. all the points inside the domain where the value of the function is greater than at neighbouring points. Finding the greatest value of a function involves taking into account its values on the *boundary* (contour) of the given domain, and comparing these values with the maxima inside the domain. The greatest of these values is *the greatest value of the function in the given domain.* The process is similar for finding the *least* value (absolute minimum) of a function in a given domain. The above remarks may be better understood with the aid of an example.

A triangle OAB (Fig. 166) *is formed on a plane by the axes OX, OY, and the line*

$$x + y - 1 = 0. \tag{16}$$

The problem is to find the point in the triangle, for which the sum of the squares of its distances from the vertices is a minimum.

On noting that the vertices A and B have coordinates (1,0) and (0,1), we can write down an expression for the above sum of the squares of the distances of a variable point (x, y) from the vertices of the triangle:

$$z = 2x^2 + 2y^2 + (x - 1)^2 + (y - 1)^2.$$

Equating to zero the first order partial derivatives gives $x = y = 1/3$, and these values are easily shown to give a minimum, with $z = 4/3$. We now consider z on the contour of the triangle. Taking z on the side OA implies putting $y = 0$:

$$z = 2x^2 + (x - 1)^2 + 1 ,$$

where x can vary in the interval (0,1). We apply the method of [60] to find that z has a least value, $z = 5/3$, at the point O on the side OA, where $x = 1/3$. Similarly, z is found to have the least value 5/3 at the point D on side OB where $y = 1/3$. We investigate the value of z on side AB by putting $y = 1 - x$, in accordance with equation (16), in the expression for z:

$$z = 3x^2 + 3(x - 1)^2 ,$$

where x varies in the interval (0,1). The least value of z here is $z = 3/2$, occurring at the point E, where $x = y = 1/2$. We thus obtain the following table of possible least values of the function:

x, y	$\frac{1}{3}, \frac{1}{3}$	$\frac{1}{3}, 0$	$0, \frac{1}{3}$	$\frac{1}{2}, \frac{1}{2}$
z	$\frac{4}{3}$	$\frac{5}{3}$	$\frac{5}{3}$	$\frac{3}{2}$

We see from the table that the least value of z, $z = 4/3$, is obtained at the point (1/3, 1/3). A solution of this problem can be obtained for any triangle, the required point being the centre of gravity of the triangle.

167. Conditional maxima and minima. We have so far considered the maxima and minima of a function, on the assumption that the variables concerned are independent variables. We now turn to functions in which the variables are connected by certain relationships, in which case, the maxima and minima are referred to as *conditional*.

Let it be required to find the maxima and minima of the function

$$f(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_{m+n})$$

of $(m + n)$ variables x_i, connected by n relationships:

$$\varphi_i(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_{m+n}) = 0 \ (i = 1, 2, \ldots, n). \qquad (17)$$

We shorten the writing in future by omitting the arguments of the functions. By solving the n relationships (17) with respect to n variables, say

$$x_{m+1}, x_{m+2}, \ldots, x_{m+n},$$

we can express these in terms of the remaining m independent variables

$$x_1, x_2, \ldots, x_m;$$

on substitution in the function f, we now get a function of m independent variables, i.e. the problem has been reduced to that of finding the ordinary maxima and minima. The proposed solution of system (17) is often laborious or indeed impossible in practice, and we indicate an alternative approach to the problem, *the method of Lagrange multipliers*.

Let the function f attain a conditional maximum, or a conditional minimum, at the point $M(x_1, x_2, \ldots, x_{m+n})$. Having assumed the existence of the derivatives at the point M, we can say that the total differential of f must vanish at M [162]:

$$\sum_{s=1}^{m+n} \frac{\partial f}{\partial x_s} dx_s = 0. \tag{18}$$

On the other hand, we obtain the following n equations at the same point M on differentiating relationships (17):

$$\sum_{s=1}^{m+n} \frac{\partial \varphi_i}{\partial x_s} = 0 \qquad (i = 1, 2, \ldots, n).$$

We multiply these latter equations by the factors

$$\lambda_1, \lambda_2, \ldots, \lambda_n,$$

that are so far undefined, and add all these term by term to relationship (18):

$$\sum_{s=1}^{m+n} \left(\frac{\partial f}{\partial x_s} + \lambda_1 \frac{\partial \varphi_1}{\partial x_s} + \lambda_2 \frac{\partial \varphi_2}{\partial x_s} + \ldots + \lambda_n \frac{\partial \varphi_n}{\partial x_s}\right) dx_s = 0. \tag{19}$$

We define these n factors so that each coefficient of the n differentials

$$dx_{m+1}, dx_{m+2}, \ldots, dx_{m+n}$$

of the dependent variables vanishes; in other words, $\lambda_1, \lambda_2, \ldots, \lambda_n$ are defined by the n equations:

$$\frac{\partial f}{\partial x_s} + \lambda_1 \frac{\partial \varphi_1}{\partial x_s} + \lambda_2 \frac{\partial \varphi_2}{\partial x_s} + \ldots + \lambda_n \frac{\partial \varphi_n}{\partial x_s} = 0 \qquad (20)$$

$$(s = m + 1,\ m + 2 \ldots,\ m + n).$$

The left-hand side of (19) now only includes terms that contain the differentials of the independent variables:

$$\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_m ,$$

i.e.

$$\sum_{s=1}^{m} \left(\frac{\partial f}{\partial x_s} + \lambda_1 \frac{\partial \varphi_1}{\partial x_s} + \lambda_2 \frac{\partial \varphi_2}{\partial x_s} + \ldots + \lambda_n \frac{\partial \varphi_n}{\partial x_s} \right) \mathrm{d}x_s = 0. \qquad (21)$$

But the differentials $\mathrm{d}x_1, \mathrm{d}x_2, \ldots, \mathrm{d}x_m$ of the independent variables are arbitrary magnitudes. We can put one of these equal to unity, and the rest zero, and it is now seen to follow from (21) that all the coefficients in the equation are zero [158], i.e.

$$\frac{\partial f}{\partial x_s} + \lambda_1 \frac{\partial \varphi_1}{\partial x_s} + \lambda_2 \frac{\partial \varphi_2}{\partial x_s} + \ldots + \lambda_n \frac{\partial \varphi_n}{\partial x_s} = 0 \qquad (22)$$

$$(s = 1, 2, \ldots, m).$$

In all the above formulae, starting with (18), the variables x_s must be assumed to be replaced by the coordinates of the point M, at which, by hypothesis, f attains a conditional maximum or minimum. This applies in particular to equations (20), from which $\lambda_1, \lambda_2, \ldots, \lambda_n$ have to be determined.

Equations (22) and (17) thus express the necessary conditions for a conditional maximum or minimum to be obtained at the point $(x_1, x_2, \ldots, x_{m+n})$.

Equations (22) *and* (17) *give us* $(m + 2n)$ *equations for finding the* $(m + n)$ *variables* x_s *and the the* n *factors* λ_i.

It is clear from system (22) *that the values of* x_s *for which the function* f *attains a conditional maximum or minimum are to be found by equating to zero the partial derivatives with respect to all* x_s *of the function* Φ, *defined by :*

$$\Phi = f + \lambda_1 \varphi_1 + \lambda_2 \varphi_2 + \ldots + \lambda_n \varphi_n ,$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ *are considered as constants, and by taking in addition the* n *equations* (17).

We deal briefly with the problem of sufficient conditions in the next section.

It is to be noted that, in deducing the above rule, we have assumed not only the existence of the derivatives of functions f and φ_i, but also the possibility of determining the factors $\lambda_1, \lambda_2, \ldots \lambda_n$ from equations (20). As a matter of fact, our rule cannot give us certain values $(x_1, x_2, \ldots, x_{m+n})$, for which a conditional maximum or minimum is obtained. This will now be explained in more detail, and the theory made more precise.

168. Supplementary remarks. Suppose that we are examining the conditional maxima and minima of a function $f(x, y)$ with the auxiliary condition:

$$\varphi(x, y) = 0\,, \tag{23}$$

and suppose that, for instance, a conditional maximum is found at the point (x_0, y_0), so that $\varphi(x_0, y_0) = 0$. Let $\varphi(x, y)$ have continuous first order derivatives at, and in the neighbourhood of, (x_0, y_0), and suppose further that

$$\varphi'_{y_0}(x_0, y_0) \neq 0\,. \tag{24}$$

Equation (23) now uniquely defines in the neighbourhood of $x = x_0$ a function $y = \omega(x)$, which is continuous, has a continuous derivative, and satisfies $y_0 = \omega(x_0)$ [157]. We can say, on substituting $y = \omega(x)$ in the function $f(x, y)$, that the function $f[x, \omega(x)]$ of a single variable x must have a maximum at $x = x_0$, and its total differential with respect to x must therefore vanish at $x = x_0$, i.e.

$$f'_{x_0}(x_0, y_0) + f'_{y_0}(x_0, y_0)\, \omega'(x_0) = 0\,.$$

Substituting $y = \omega(x)$ in (23) and differentiating with respect to x gives us at (x_0, y_0) [69]:

$$\varphi'_{x_0}(x_0, y_0) + \varphi'_{y_0}(x_0, y_0)\, \omega'(x_0) = 0\,.$$

We obtain, on multiplying the second equation by λ and adding term by term to the first:

$$(f'_{x_0} + \lambda\varphi'_{x_0}) + (f'_{y_0} + \lambda\varphi'_{y_0})\, \omega'(x_0) = 0\,.$$

On defining λ by the condition $f'_{y_0} + \lambda\varphi'_{y_0} = 0$, as is possible by (24), we get $f'_{x_0} + \lambda\varphi'_{x_0} = 0$, i.e. we arrive at the two equations:

$$f'_{x_0} + \lambda\varphi'_{x_0} = 0\,; \quad f'_{y_0} + \lambda\varphi'_{y_0} = 0\,, \tag{25}$$

with which the equation $\varphi(x_0, y_0) = 0$ is to be associated; this in fact verifies the multiplier method. If condition (24) is not satisfied, i.e. $\varphi'_{y_0}(x_0, y_0) = 0$, but on the other hand $\varphi'_{x_0}(x_0, y_0) \neq 0$, all the above argument can be repeated, with the roles of x and y interchanged. If we have at (x_0, y_0):

$$\varphi'_{x_0}(x_0, y_0) = 0 \text{ and } \varphi'_{y_0}(x_0, y_0) = 0\,, \tag{26}$$

we cannot prove that the point (x_0, y_0) is obtained by means of the method of multipliers.

Equations (26) show that (x_0, y_0) is a singular point of curve (23) [76]. We now give an example of a problem in which condition (26) holds at a conditional minimum.

Let it be required to find the shortest distance from the point $(-1, 0)$ to points lying on the semi-cubical parabola $y^2 - x^3 = 0$, illustrated in Fig. 87 [76]. The minimum is thus required of the function $f = (x+1)^2 + y^2$, with the additional condition $\varphi = y^2 - x^3 = 0$. It is clear geometrically that the minimum is found at the point (0,0) of the parabola, which is a singular point of the curve. The method of multipliers leads us to the following two equations:

$$2(x+1) - 3\lambda x^2 = 0, \ 2y + 2\lambda y = 0.$$

On substituting $x = 0$, $y = 0$, the first equation leads to the absurdity $2 = 0$, whilst the second is satisfied for all λ. The method of multipliers does not lead us in this case to the point (0,0) at which the conditional minimum is reached. It can be shown similarly that, if a function has a maximum or a minimum at a point (x_0, y_0, z_0), where there is the additional condition $\varphi(x, y, z) = 0$, and where at least one of the first order partial derivatives of φ differs from zero at (x_0, y_0, z_0), this point can now be obtained by the method of multipliers.

Similar arguments apply in more general cases; but we have now reached the point of referring to the theorem on the existence of implicit functions for a system of equations, which we mentioned in [157]. Let the function $f(x, y, z)$, for example, have a conditional maximum at the point (x_0, y_0, z_0), with the two additional conditions

$$\varphi(x, y, z) = 0, \ \psi(x, y, z) = 0\,, \tag{27}$$

and with the usual assumptions regarding the existence and continuity of the derivatives; and let us also have

$$\varphi'_{y_0}(x_0, y_0, z_0)\, \psi'_{z_0}(x_0, y_0, z_0) - \varphi'_{z_0}(x_0, y_0, z_0)\, \psi'_{y_0}(x_0, y_0, z_0) \neq 0. \tag{28}$$

Equations (27) now uniquely define the functions: $y = \omega_1(x)$, $z = \omega_2(x)$, such that $y_0 = \omega_1(x_0)$; $z_0 = \omega_2(x_0)$. Substitution of these functions in f gives us a function of x only, with a maximum at $x = x_0$, whence:

$$f'_{x_0}(x_0, y_0, z_0) + f'_{y_0}(x_0, y_0, z_0)\, \omega'_1(x_0) + f'_{z_0}(x_0, y_0, z_0)\, \omega'_2(x_0) = 0.$$

On substituting the same functions in (27) and differentiating with respect to x at the point (x_0, y_0, z_0), we get:

$$\varphi'_{x_0} + \varphi'_{y_0}\omega'_1(x_0) + \varphi'_{z_0}\omega'_2(z_0) = 0\,; \quad \psi'_{x_0} + \psi'_{y_0}\omega'_1(x_0) + \psi'_{z_0}\omega'_2(x_0) = 0\,.$$

We multiply these equations by λ_1, λ_2 and add to the previous equation:

$$(f'_{x_0} + \lambda_1\varphi'_{x_0} + \lambda_2\psi'_{x_0}) + (f'_{y_0} + \lambda_1\varphi'_{y_0} + \lambda_2\psi'_{y_0}) + (f'_{z_0} + \lambda_1\varphi'_{z_0} + \lambda_2\psi'_{z_0}) = 0\,. \tag{29}$$

We can assert, on taking (28) into account, that λ_1 and λ_2 can be defined from the two equations

$$f'_{y_0} + \lambda_1\varphi'_{y_0} + \lambda_2\psi'_{y_0} = 0\,; \quad f'_{z_0} + \lambda_1\varphi'_{z_0} + \lambda_2\psi'_{z_0} = 0\,, \tag{30}$$

after which, (29) reduces to

$$f'_{x_0} + \lambda_1 \varphi'_{x_0} + \lambda_2 \psi'_{x_0} = 0\,, \tag{31}$$

which verifies the method of multipliers in the present case. We must also take, *in addition* to equations (30) and (31):

$$\varphi(x_0, y_0, z_0) = 0 \text{ and } \psi(x_0, y_0, z_0) = 0.$$

We might have replaced (28) by a similar condition, obtained by differentiating with respect to x_0 and y_0 or x_0 and z_0 instead of with respect to y_0 and z_0. But if the expression on the left of (28) is zero, and the analogous expressions, obtained by differentiating with respect to x_0 and y_0 or x_0 and z_0, are also zero, we are unable to justify the method of multipliers for the point (x_0, y_0, z_0). It can be shown that this special case cannot arise in any of the examples of the next section. For example, we have one auxiliary condition (32) in example 1, and at least one of the numbers A, B and C differs from zero on the left-hand side of the condition. If $C \neq 0$, for instance, the derivative of the left of (32) with respect to z is equal to C, and therefore differs from zero at every point (x, y, z). This indicates that an answer must be obtainable here by the method of multipliers.

We now briefly discuss the sufficient conditions for a conditional maximum or minimum, whilst confining ourselves to the case of two independent variables. Let us look for the conditional maxima and minima of a function $f(x,y,z)$, with the condition $\varphi(x, y, z) = 0$. We form a new function $\Phi = f + \lambda\varphi$. We equate the first order partial derivatives with respect to x, y, z to zero and solve the equations with the added condition, and we thus obtain, say, $x = x_0, y = y_0, z = z_0, \lambda = \lambda_0$. We have to investigate the values obtained, i.e. find the sign of the difference $f(x, y, z) - f(x_0, y_0, z_0)$ for all (x, y, z) sufficiently near (x_0, y_0, z_0) and satisfying our condition $\varphi(x, y, z) = 0$. We introduce the function $\psi(x, y, z) = f(x, y, z) + \lambda_0\, \varphi(x, y, z)$. It follows immediately from the equation of the condition that we can take the difference $\psi(x, y, z) - \psi(x_0, y_0, z_0)$, and investigate its sign, instead of taking $f(x, y, z) - f(x_0, y_0, z_0)$. The first order partial derivatives of ψ vanish at (x_0, y_0, z_0) by hypothesis. On expanding the latter difference by Taylor's formula and confining ourselves to the second derivatives, we get an expression of the form [cf. 165]:

$$\psi(x, y, z) - \psi(x_0, y_0, z_0) = a_{11}\, dx^2 + a_{22}\, dy^2 + a_{33}\, dz^2 + 2a_{12}\, dxdy +$$

$$+ 2a_1a_3\, dxdz + 2a_{23}\, dydz + \dots ,$$

where a_{ik} denote the values of the corresponding second order partial derivatives of $\psi(x, y, z)$ at (x_0, y_0, z_0), and dx, dy, dz denote the increments of the variables. We take $\varphi_{z_0}(x_0, y_0, z_0) \neq 0$, so that the equation of the condition defines $z = \omega(x, y)$, where $z_0 = \omega(x_0\, y_0)$. We get from the equation of the condition:

$$\varphi_x(x, y, z)\, dx + \varphi_y(x, y, z)\, dy + \varphi_z(x, y\, z)\, dz = 0.$$

On substituting here $x = x_0$, $y = y_0$, $z = z_0$, we can express dz in terms of dx and dy:

$$dz = \frac{\varphi_{x_0}(x_0, y_0, z_0)}{\varphi_{z_0}(x_0, y_0, z_0)} dx - \frac{\varphi_{y_0}(x_0, y_0, z_0)}{\varphi_{z_0}(x_0, y_0, z_0)} dy.$$

We substitute the expression for dz in the previous formula and collect similar terms, and obtain:

$$\psi(x, y, z) - \psi(x_0, y_0, z_0) = A dx^2 + 2B dx dy + C dy^2 + \ldots$$

We can now use the test for maxima and minima of [163]. For instance, if $AC - B^2 > 0$ and $A > 0$, $f(x, y, z)$ has a conditional minimum at (x_0, y_0, z_0). It immediately follows from the discussion of [163] that a sufficient basis for the rule concerned is the assumption that $f(x, y, z)$ and $\varphi(x, y, z)$ have continuous derivatives up to the second order at, and in the neighbourhood of, the point (x_0, y_0, z_0).

We shall not dwell further on the question of sufficient conditions for conditional maxima and minima. The essential points in the above argument were the replacement of the difference $f(x, y, z) - f(x_0, y_0, z_0)$ by $\psi(x, y, z) - \psi(x_0, y_0, z_0)$, the first order derivatives of which vanish at (x_0, y_0, z_0), and the fact that the differential dz of the dependent variable was defined in terms of the differentials dx, dy of the independent variables by an equation of the first degree. A similar approach must be used when considering the sufficient conditions for any other number of variables and conditions.

169. Examples. 1. *We are required to find the shortest distance from the point* (a, b, c) *to the plane*

$$Ax + By + Cz + D = 0. \tag{32}$$

The square of the distance from the point (a, b, c) to the variable point (x, y, z) is given by

$$r^2 = (x - a)^2 + (y - b)^2 + (z - c)^2. \tag{33}$$

The coordinates (x, y, z) must here satisfy equation (32) (the point must lie on the plane). We find the minimum of expression (33), with the condition (32). We form the function:

$$\Phi = (x - a)^2 + (y - b)^2 + (z - c)^2 + \lambda_1(Ax + By + Cz + D).$$

Equating to zero its partial derivatives with respect to x, y, z gives:

$$x = a - \frac{1}{2}\lambda_1 A, \quad y = b - \frac{1}{2}\lambda_1 B, \quad z = c - \frac{1}{2}\lambda_1 C. \tag{34}$$

We can find λ_1 by substituting these expressions in (32):

$$\lambda_1 = \frac{2(Aa + Bb + Cc + D)}{A^2 + B^2 + C^2}. \tag{35}$$

We have obtained unique values for the variables, and these must give the required least value, since this necessarily exists. Substitution of the

values of (34) in (33) gives us the expression for the square of the distance from the point to the plane:

$$r_0^2 = \frac{1}{4} \lambda_1^2 (A^2 + B^2 + C^2),$$

where λ_1 is defined by (35).

2. *To express a positive number a as the sum of three positive terms x, y, z, so as to obtain the greatest value of*

$$x^m y^n z^p \tag{36}$$

(m, n, p are given positive numbers). We have to find the maximum of expression (36) with the condition

$$x + y + z = a. \tag{37}$$

Instead of looking for the maximum of (36), we can look for the maximum of its logarithm

$$m \log x + n \log y + p \log z.$$

We form the function:

$$\Phi = m \log x + n \log y + p \log z + \lambda_1(x + y + z - a).$$

We obtain, on equating the partial derivatives to zero:

$$x = -\frac{m}{\lambda_1}, \quad y = -\frac{n}{\lambda_1}, \quad z = -\frac{p}{\lambda_1},$$

and (37) now gives:

$$\lambda_1 = -\frac{m + n + p}{a},$$

i.e.

$$x = \frac{ma}{m + n + p}, \quad y = \frac{na}{m + n + p}, \quad z = \frac{pa}{m + n + p}, \tag{38}$$

the values obtained for the variables being positive numbers. It can be shown that (36) must have a greatest value with the conditions postulated, and the uniqueness of the result shows, as in Example 1, that this greatest value must in fact be given by the values obtained for the variables.

It can be seen from (38) that a has to be split into components proportional to the exponents m, n, p, to obtain the solution of the problem.

We suggest that the reader carry out an investigation of the sufficient conditions in the last two examples, by the method outlined in the previous section.

3. *A conductor of length l_0 branches at one end into k separate conductors of lengths l_s ($s = 1, 2, \ldots, k$), the corresponding currents in the individual conductors being $i_0, i_1, \ldots, i_k$. The problem is to choose the cross-sectional areas $q_0, q_1, \ldots, q_k$ of the conductors so as to employ the least quantity of material*

V for a given potential difference over the circuits $(l_0, l_1), (l_0, l_2), \ldots, (l_0, l_k)$. (Fig. 167)

Let c denote the resistance of a wire of the given material, of unit length and unit cross-sectional area.

We require the least value of the function V of variables $q_0, q_1, \ldots, q_k$, given by

$$V = l_0 q_0 + l_1 q_1 + \ldots + l_k q_k.$$

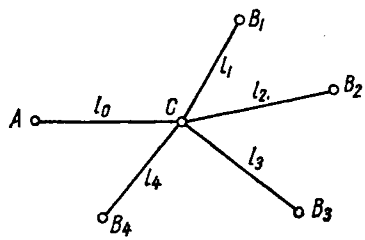

FIG. 167

We can use the given potential difference E to write down k relationships:

$$\varphi_s = c\left(\frac{l_0 i_0}{q_0} + \frac{l_s i_s}{q_s}\right) - E = 0 \quad (s = 1, 2, \ldots, k). \tag{39}$$

We form the function:

$$\Phi = (l_0 q_0 + l_1 q_1 + \ldots + l_k q_k) + \sum_{s=1}^{k} \lambda_s \left[c\left(\frac{l_0 i_0}{q_0} + \frac{l_s i_s}{q_s}\right) - E\right].$$

We equate to zero the partial derivatives of Φ with respect to $q_0, q_1, \ldots, q_k$:

$$\left.\begin{aligned} &l_0 - \frac{c l_0 i_0}{q_0^2}(\lambda_1 + \lambda_2 + \ldots + \lambda_k) = 0, \\ &l_s - \frac{\lambda_s c l_s i_s}{q_s^2} = 0 \quad (s = 1, 2, \ldots, k). \end{aligned}\right\} \tag{40}$$

We have from condition (39):

$$\frac{l_1 i_1}{q_1} = \frac{l_2 i_2}{q_2} = \ldots = \frac{l_k i_k}{q_k} = \frac{E}{c} - \frac{l_0 i_0}{q_0};$$

if we let σ denote the common value of these ratios, we can write:

$$q_s = \frac{l_s i_s}{\sigma} \quad (s = 1, 2, \ldots, k), \qquad \sigma = \frac{E}{c} - \frac{l_0 i_0}{q_0}. \tag{41}$$

We obtain from equations (40):

$$\lambda_s = \frac{q_s^2}{c i_s} = \frac{l_s^2 i_s}{c\sigma^2}.$$

On substituting these expressions for λ_s in the first of equations (40), we get:

$$q_0^2 = \frac{i_0}{\sigma^2}(l_1^2 i_1 + l_2^2 i_2 + \ldots + l_k^2 i_k),$$

or

$$q_0 = \frac{\sqrt{i_0(l_1^2 i_1 + l_2^2 i_2 + \ldots + l_k^2 i_k)}}{\dfrac{E}{c} - \dfrac{l_0 i_0}{q_0}},$$

whence finally:

$$q_0 = \frac{c}{E}\left[l_0 i_0 + \sqrt{i_0(l_1^2 i_1 + l_2^2 i_2 + \ldots + l_k^2 i_k)}\right].$$

We substitute this expression for q_0 in (41) and obtain for $q_1, q_2, \ldots, q_k$:

$$q_s = \frac{c l_s i_s}{E}\left(1 + \frac{l_0 i_0}{\sqrt{i_0(l_1^2 i_1 + l_2^2 i_2 + \ldots + l_k^2 i_k)}}\right) \qquad (s = 1, 2, \ldots, k).$$

The necessary conditions for a maximum or minimum of V are seen to give us a unique set of positive values for $q_0, q_1, \ldots, q_k$, and these must represent the solution of our problem, since it is evident from physical considerations that a minimum quantity of material must be obtained for a certain choice of the cross-sectional areas.

EXERCISES ON CHAPTER V

1. Find $f(1/2, 3), f(1, -1)$ for $f(x, y) = xy + x/y$.

2. Find $f(y, x), f(-x, -y), f(x^{-1}, y^{-1}), 1/f(x, y)$ for $f(x, y) = \left(\dfrac{x^2 - y^2}{2xy}\right)$.

3. Find the value of $f(x, y) = 1 + x - y$ at points on the parabola $y = x^2$.

4. Find the value of $z = (x^4 + 2x^2 y^2 + y^4)/(1 - x^2 - y^2)$ at points on the circle $x^2 + y^2 = R^2$.

5. Define $f(x)$ if $f(y/x) = y^{-1}(x^2 + y^2)^{1/2}$, $y > 0$.

6. Find $f(x, y)$ if $f(x + y, x - y) = xy + y^2$.

7. Let $z = \sqrt{y} + f(\sqrt{x} - 1)$. Determine the functions f and z if $z = x$ when $y = 1$.

8. Let $z = xf(y/x)$. Determine the functions f and z if $z = \sqrt{(1 + y^2)}$ when $x = 1$.

9. Find and represent the domains of existence of the following functions:

(a) $z = \sqrt{(1 - x^2 - y^2)}$; (b) $z = 1 + \sqrt{-(x - y)^2}$; (c) $z = \log(x + y)$;
(d) $z = x + \arccos y$; (e) $z = \sqrt{(1 - x^2)} + \sqrt{(1 - y^2)}$;
(f) $z = \arcsin(y/x)$; (g) $z = \sqrt{(x^2 - 4)} + \sqrt{(4 - y^2)}$;
(h) $z = \sqrt{[(x^2 + y^2 - a^2)(2a^2 - x^2 - y^2)]}$ $(a > 0)$;

(i) $z = \sqrt{(y \sin x)}$; (j) $z = \log(x^2 + y)$;

(k) $z = \arctan \dfrac{x - y}{1 + x^2 y^2}$;

(l) $z = (x^2 + y^2)^{-1}$; (m) $z = (y - x^{1/2})^{-1/2}$;

(n) $z = \dfrac{1}{x-1} + \dfrac{1}{y}$; (o) $z = \sqrt{[\sin(x^2 + y^2)]}$.

Find the partial derivatives of the functions **10–15**:

10. $u = (xy)^z$. **11.** $u = z^{xy}$. **12.** $u = x^3 + y^3 + z^3 - 3xyz$.

13. $u = x^2y^2 t/(1 - z^2)$. **14.** $u = r^{-1}$, where $r = \sqrt{(x^2 + y^2 + z^2)}$.

15. $u = xyz/(x + y)(z + 1)$

16. Find f_x, f_y, f_z at the point (1, 2, 0) for the function $f(x, y, z) = \log(xy + z)$.

17. Calculate the value of the determinant

$$\begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} & \dfrac{\partial x}{\partial \varphi} \\[2ex] \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} & \dfrac{\partial y}{\partial \varphi} \\[2ex] \dfrac{\partial z}{\partial r} & \dfrac{\partial z}{\partial \theta} & \dfrac{\partial z}{\partial \varphi} \end{vmatrix}$$

for the functions $x = r \cos \varphi \sin \theta$, $y = r \sin \varphi \sin \theta$, $z = r \cos \theta$.

Verify Euler's theorem (on homogeneous functions) for the functions **18–23**:

18. $z = ax^2 + 2hxy + by^2$. **19.** $z = x/(x^2 + y^2)$.

20. $z = (x + y)(x^2 + y^2)^{1/3}$. **21.** $z = \log(y/x)$.

22. $z = \log(x + y) - \log(x + z)$. **23.** $z = xy \sin(x/y)$.

24. Show that if $z = \log(x^2 + xy + y^2)$,

$$x \frac{\partial z}{\partial x} + y \frac{\partial z}{\partial y} = 2.$$

25. Show that if $z = xy + xe^{y/x}$,

$$x \frac{\partial z}{\partial x} + y \frac{\partial z}{\partial y} = xy + z.$$

26. Show that if $u = (x - y)(y - z)(z - x)$, then

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z} = 0.$$

27. Show that if $u = x + (x - y)/(y - z)$,

$$\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y} + \frac{\partial u}{\partial z} = 1.$$

28. Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{dz}{dx}$ if $z = \arctan \dfrac{y}{x}$ and $y = x^2$.

29. Find $\dfrac{\partial z}{\partial x}$ and $\dfrac{dz}{dx}$ if $z = x^y$ where $y = \varphi(x)$.

30. Find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ if $z = f(u, v)$, where $u = x^2 - y^2$, $v = e^{xy}$.

31. Find $\frac{\partial z}{\partial u}$ and $\frac{\partial z}{\partial v}$ if $z = \arctan \frac{y}{x}$, where $x = u \sin v$, $y = u \cos v$.

32. Find $\frac{\partial z}{\partial x}$ and $\frac{\partial z}{\partial y}$ if $z = f(u)$, where $u = xy + y/x$.

33. Prove that if $u = \Phi(x^2 + y^2 + z^2)$, where $x = r \cos \varphi \sin \theta$, $y = r \sin \varphi \sin \theta$, $z = r \cos \theta$, then

$$\frac{\partial u}{\partial \theta} = \frac{\partial u}{\partial \varphi} = 0.$$

34. Prove that if $z = f(x + ay)$, where f is a differentiable function, then

$$\frac{\partial z}{\partial y} = a \frac{\partial z}{\partial x}.$$

35. Prove that the function $w = f(u, v)$ where $u = x + at$, $v = y + bt$, satisfies the equation

$$\frac{\partial w}{\partial t} = a \frac{\partial w}{\partial x} + b \frac{\partial w}{\partial y}.$$

36. Find $\frac{\partial^2 z}{\partial x^2}$, $\frac{\partial^2 z}{\partial x \, \partial y}$, $\frac{\partial^2 z}{\partial y^2}$ if $z = c(x^2/a^2 + y^2/b^2)^{1/2}$.

37. Find $\frac{\partial^2 z}{\partial x^2}$, $\frac{\partial^2 z}{\partial x \, \partial y}$, $\frac{\partial^2 z}{\partial y^2}$ if $z = \log(x^2 + y)$.

38. Find $\frac{\partial^2 z}{\partial x \, \partial y}$ if $z = \sqrt{2xy + y^2}$.

39. Prove that for the function $z = \arcsin \sqrt{(x - y)/z}$, $\frac{\partial^2 z}{\partial x \, \partial y} = \frac{\partial^2 z}{\partial y \, \partial x}$.

40. Prove that for the function $z = x^y$, $\frac{\partial^2 z}{\partial x \, \partial y} = \frac{\partial^2 z}{\partial y \, \partial x}$.

41. Prove that for the function defined by the equation

$$f(x, y) = \begin{cases} xy\,(x^2 - y^2)/(x^2 + y^2) & \text{if } (x, y) \neq (0, 0), \\ 0 & \text{at } (0, 0), \end{cases}$$

$$f_{xy}(0, 0) = -1, \qquad f_{yx}(0, 0) = +1.$$

42. Prove that the function $u = \log r$, where $r = \sqrt{[(x - a)^2 + (y - b)^2]}$ satisfies the two-dimensional Laplace equation

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0.$$

43. Prove the function

$$u(x, t) = A \sin(\lambda\, at + \Phi) \sin \lambda\, x\,,$$

where λ, Φ, a are constants, satisfies the one-dimensional wave equation

$$\frac{\partial^2 u}{\partial t^2} = a^2 \frac{\partial^2 u}{\partial y^2}.$$

Investigate the maxima and minima of the functions **44–52**:

44. $(x-1)^2+2y^2$. **45.** $(x-1)^2-2y^2$. **46.** x^2+xy+y^2-2x-y.

47. $x^3y^2(6-x-y), x>0, y>0$. **48.** $x^4+y^4-2x^2+4xy-2y^2$.

49. $xy\sqrt{1-x^2/a^2-y^2/b^2}$. **50.** $1-(x^2+y^2)^{2/3}$.

51. $(x^2+y^2)\,e^{-(x^2+y^2)}$. **52.** $(1+x-y)(1+x^2+y^2)^{-1/2}$.

53. Find the turning values of the functions defined by the equations

(*a*) $x^2+y^2+z^2-2x+4y-6z-11=0$;

(*b*) $x^3-y^2-3x+4y+z^2+z-8=0$.

Determine the conditional extrema of the functions **54—59**:

54. $z=xy$ with $x+y=1$. **55.** $z=x+2y$ with $x^2+y^2=5$.

56. $z=x^2+y^2$ with $\frac{1}{2}x+\frac{1}{3}y=1$.

57. $z=\cos^2 x+\cos^2 y$ with $y-x=\frac{1}{4}\pi$.

58. $u=x-2y+22$ with $x^2+y^2+z^2=9$.

59. $u=x^2+y^2+z^2$ with $\frac{x^2}{a^2}+\frac{y^2}{b^2}+\frac{z^2}{c^2}=1, (a>b>c>0)$.

60. From all rectangular parallelepipeds of prescribed volume V, find the one which has minimum surface area.

61. If an open rectangular box has prescribed volume V, find the dimensions if its surface area is to be the least possible.

62. Find the shape of the triangle with prescribed perimeter $2p$ enclosing the maximum area.

63. Find the shape of the rectangular parallepiped of prescribed surface area S enclosing the maximum volume.

64. Find the triangle of prescribed perimeter $2p$, which when rotated about one of its sides generates a solid with the greatest possible volume.

65. Masses of amount m_1, m_2, m_3 are placed at the points $P_1(x_1, y_1)$, $P_2(x_2, y_2)$, $P_3(x_3, y_3)$. Find the position of the point P such that the moment of inertia of these masses relative to P is a minimum.

66. Find the equation of the plane through the point (a, b, c) which together with the coordinate planes forms a tetrahedron of minimum volume.

67. Find the dimensions of the rectangular parallelepiped of maximum volume which can be inscribed in an ellipsoid.

CHAPTER VI

COMPLEX NUMBERS. THE FOUNDATIONS OF HIGHER ALGEBRA. INTEGRATION OF VARIOUS FUNCTIONS

§ 17. Complex numbers

170. Complex numbers. As we know, the operation of extracting a root is not always possible if we confine ourselves to real numbers; the root of even degree of a negative number has no meaning in the domain of real numbers; and moreover, a quadratic equation with real coefficients does not always possess real roots. This situation leads naturally to a broadening of the concept of number by means of the introduction of new numbers of a more general kind, with real numbers appearing as a particular case. These new numbers, and the operations on them, must be defined in such a way that all the basic arithmetical laws that are known for real numbers remain valid. The fact that such a definition is possible is established below.

Apart from the impossibility just mentioned of extracting a root in certain cases, simple geometrical considerations also lead to a natural extension of the concept of number. It is the geometrical approach to the problem that we shall choose.

We know that every real number can be represented graphically either by a segment, cut off from a given axis OX, or else by a point on this axis, if we agree to locate the initial point of all segments at the origin of coordinates. Conversely, a definite real number corresponds to every segment or point on the axis OX.

If we now consider the whole *plane*, in reference to the coordinate axes OX, OY, instead of taking the single axis OX, suitable generalization of the concept of number will enable us to associate a certain number, which we call *complex*, with every *vector*, or with every *point*, lying on this plane.

If we agree not to distinguish between vectors of the same length and the same direction, a real number can be associated not only with every vector along OX, but with every vector in general that is parallel to OX. The real number one, in particular, corresponds to a vector of unit length in the positive direction of OX.

We associate the symbol i, called *imaginary unity*, with a vector of unit length in the positive direction of the axis OY. Every vector $\overline{MN}$ in the plane can be represented as the sum of two vectors $\overline{MP}$ and $\overline{PN}$, parallel to the coordinate axes (Fig. 168). Some real number a corresponds to the vector $\overline{MP}$, parallel to OX. Let the symbol bi correspond to the vector $\overline{PN}$, parallel to OY, where b is a real number which is equal in absolute value to the length of $\overline{PN}$, and which is positive or negative according as the direction of PN coincides with the positive or negative direction of OY. We now naturally associate with the vector $\overline{MN}$ a *complex number*, of the form

$$a + bi.$$

We note the fact that the $(+)$ sign in the expression written, $a + bi$, is not the sign of an operation. The expression is to be viewed as a single symbol for denoting the complex number. We shall give our attention to the sign after defining addition of complex numbers.

The real numbers a and b evidently consist of the magnitudes of the projections of the vector $\overline{MN}$ on the coordinate axes.

We mark off the vector $\overline{OA}$ from the origin of coordinates (Fig. 168), with the same length and direction as $\overline{MN}$. The end A of the vector will have coordinates (a, b), and we can associate the same complex number $a + bi$ with the point A as with vectors $\overline{MN}$ and $\overline{OA}$.

Thus, *a definite complex number $a + bi$ corresponds to every vector in the plane (to every point of the plane). The real numbers a and b are equal to the projections of the vector on the coordinate axes (to the coordinates of the point).*

We obtain the set of complex numbers on assigning all possible real values to the symbols a and b in the expression $a + bi$. We call a the *real part*, and b the *imaginary part*, of the complex number.

In the particular case of a vector parallel to OX, the complex number coincides with its real part:

$$a + 0i = a. \tag{1}$$

We look on the real number a as a particular case of a complex number, in accordance with formula (1).

The concept of the equality of two complex numbers follows from the geometrical interpretation. Two vectors are looked on as equal if they have the same length and the same direction, i.e. if they have the same projections on the coordinate axes; hence, *two complex numbers are considered to be equal when, and only when, their real parts and their imaginary parts are separately equal, the condition for equality of complex numbers being :*

$$a_1 + b_1 i = a_2 + b_2 i \text{ implies } a_1 = a_2, b_1 = b_2. \qquad (2)$$

In particular, $a + bi = 0$ implies $a = 0$, $b = 0$.

Instead of defining the vector $\overline{MN}$ by its projections a and b on the axes, we can use two other magnitudes: its length r and the angle φ that the direction $\overline{MN}$ makes with the positive direction of OX (Fig. 168). If we take the complex number $a + bi$ as corresponding to the point with coordinates (a, b), r and φ are evidently the polar coordinates of the point. We know that

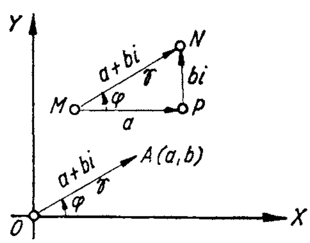

Fig. 168

$$a = r\cos\varphi;\; b = r\sin\varphi;$$

$$\left.\begin{aligned} r = \sqrt{a^2+b^2};\quad \cos\varphi = \frac{a}{\sqrt{a^2+b^2}};\quad \sin\varphi = \frac{b}{\sqrt{a^2+b^2}};\\ \varphi = \operatorname{arc\,tan}\frac{b}{a}. \end{aligned}\right\} \qquad (3)$$

The positive number r is called the modulus, and φ the amplitude, of the complex number $a + bi$. The amplitude is only defined to an accuracy of plus or minus a multiple of 2π, since any vector $\overline{MN}$ remains unchanged on rotation any integral number of times in either direction about the point M. When $r = 0$, the complex number is equal to zero, and its amplitude is completely indeterminate. The *condition for equality* of two complex numbers evidently consists in the fact that *their moduli must be equal, and their amplitudes must only differ by some multiple of* 2π.

A real number has the amplitude $2k\pi$ if it is positive, and $(2k+1)\pi$ if it is negative, where k is any integer. If the real part of a complex number is zero, the complex number has the form bi and is said to be *pure imaginary*. The vector corresponding to such a number is if parallel to OY, whilst the amplitude of bi is equal to $(\pi/2 + 2k\pi)$ $b > 0$, and $(3\pi/2 + 2k\pi)$ if $b < 0$.

The modulus of a real number is the same as its absolute value. The modulus of the number $a + bi$ is denoted by writing the number between two vertical strokes:

$$|a + bi| = \sqrt{a^2 + b^2}.$$

In future, we shall often denote a complex number by a single letter. If a is the complex number, its modulus is denoted by $|a|$. We can utilize expressions (3) for a and b to express a complex number in terms of its modulus and amplitude, in the form:

$$r(\cos\varphi + i\sin\varphi).$$

The complex number is said to be expressed in trigonometric form in this case.

171. Addition and subtraction of complex numbers. A vector sum consists of the closing side of a polygon made up of the added vectors. Since the projection of the closing side is equal to the sum of the projections of the components, we arrive at the following *definition of addition of complex numbers :*

$$(a_1 + b_1 i) + (a_2 + b_2 i) + \ldots + (a_n + b_n i) =$$
$$= (a_1 + a_2 + \ldots + a_n) + (b_1 + b_2 + \ldots + b_n)\, i. \qquad (4)$$

It is easily shown that a sum of complex numbers is independent of the order of the terms (law of transposition), and that the terms can be combined in groups (law of association), since these properties belong to the sums of the real numbers a_k and b_k.

As was mentioned above, the complex number $a + 0i$ is identical with the real number a. Similarly, the number $0 + bi$ can simply be written as bi (a pure imaginary number). We can use the definition of addition to assert that the complex number $a + bi$ is the sum of the real number a and the pure imaginary number bi, i.e. $a + bi = (a + 0i) + (0 + bi)$.

Subtraction is defined as the converse operation to addition, i.e. the difference

$$x + yi = (a_1 + b_1 i) - (a_2 + b_2 i)$$

is defined by the condition

$$(x + yi) + (a_2 + b_2 i) = a_1 + b_1 i ,$$

or, by (4) and (2): $x + a_2 = a_1$; $y + b_2 = b_1$, i.e. $x = a_1 - a_2$, $y = b_1 - b_2$, which finally gives

$$(a_1 + b_1 i) - (a_2 + b_2 i) = (a_1 - a_2) + (b_1 - b_2) i . \tag{5}$$

Subtraction of the complex number $(a_2 + b_2 i)$ from $(a_1 + b_1 i)$ is seen to be equivalent to addition to the latter of the complex number

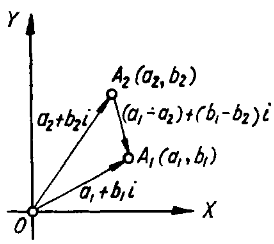

FIG. 169

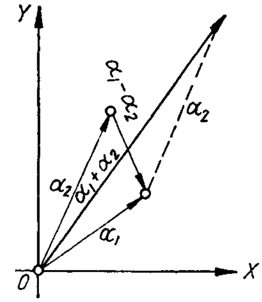

FIG. 170

$(-a_2 - b_2 i)$. This corresponds to the following: *subtraction of a vector from a second vector is equivalent to the addition to the second vector of a vector of equal magnitude but opposite direction to the first.*

We consider the vector $\overline{A_2 A_1}$, with its initial point A_2 corresponding to the complex number $a_2 + b_2 i$ and its terminal point A_1 corresponding to $a_1 + b_1 i$. This vector evidently consists of the difference of vectors $\overline{OA_1}$ and $\overline{OA_2}$ (Fig. 169), and its corresponding complex number is therefore

$$(a_1 - a_2) + (b_1 - b_2) i ,$$

equal to the difference between the complex numbers corresponding to its initial and terminal points.

We now establish the properties of the moduli of the sum and difference of two complex numbers. Since the modulus of a complex number is equal to the length of the corresponding vector, whilst one side of a triangle is less than the sum of the other two sides, we have (Fig. 170):

$$| \alpha_1 + \alpha_2 | \leqslant | \alpha_1 | + | \alpha_2 | ,$$

where the sign of equality only occurs when the vectors corresponding to a_1 and a_2 have the same direction, i.e. when the numbers have amplitudes that are either equal or differ by a multiple of 2π. This property is evidently valid for any number of terms:

$$|a_1 + a_2 + \ldots + a_n| \leqslant |a_1| + |a_2| + \ldots + |a_n|,$$

i.e. *the modulus of a sum is less than or equal to the sum of the moduli of the terms, the sign of equality only occurring when the amplitudes of the terms are equal or differ by a multiple of* 2π.

Since one side of a triangle is greater than the difference of the two remaining sides, we can also write:

$$|a_1 + a_2| \geqslant |a_1| - |a_2|,$$

i.e. *the modulus of the sum of two terms is greater than or equal to the difference of the moduli of the terms.* Equality will only occur when the corresponding vectors have opposite directions.

The subtraction of vectors and complex numbers leads to an addition, as we saw above, and we have for the modulus of the difference of two complex numbers, as for the modulus of the sum (Fig. 170):

$$|a_1| - |a_2| \leqslant |a_1 - a_2| \leqslant |a_1| + |a_2|.$$

172. Multiplication of complex numbers. The definition of the product of two complex numbers is similar to the definition of the product of two real numbers, which is: the product is taken to be the number composed of the multiplicand, as the multiplier is composed of unity. The vector corresponding to the complex number with modulus r and amplitude φ can be obtained from the *unit vector*, with unit length and directed in the positive direction of OX, by increasing its length r times and rotating it positively through an angle φ.

The product of a vector a_1 and a vector a_2 is defined as the vector obtained by applying the same elongation and the same rotation to a_1 as are applied to the unit vector in order to obtain a_2. The unit vector here obviously corresponds to real unity.

If (r_1, φ_1), (r_2, φ_2) are the moduli and amplitudes of the complex numbers corresponding to vectors a_1 and a_2, the complex number corresponding to the product of these vectors will clearly have modulus $r_1 r_2$ and amplitude $(\varphi_1 + \varphi_2)$. We thus arrive at the following definition of the product of complex numbers:

The product of two complex numbers is defined as the complex number the modulus of which is equal to the product of the moduli of the factors and the amplitude of which is the sum of the amplitudes of the factors.

When the complex numbers are given in trigonometric form, we have:

$$r_1(\cos\varphi_1 + i\sin\varphi_1)\cdot r_2(\cos\varphi_2 + i\sin\varphi_2) =$$
$$= r_1 r_2[\cos(\varphi_1+\varphi_2) + i\sin(\varphi_1+\varphi_2)]. \tag{6}$$

We now deduce the rule for finding the product of two complex numbers that are not given in trigonometric form:

$$(a_1 + b_1 i)(a_2 + b_2 i) = x + yi.$$

We can write, using the above notation for the moduli and the amplitudes of the factors:

$$a_1 = r_1\cos\varphi_1;\ b_1 = r_1\sin\varphi_1;\ a_2 = r_2\cos\varphi_2;\ b_2 = r_2\sin\varphi_2,$$

and in accordance with the definition (6) of the product:

$$x = r_1 r_2\cos(\varphi_1+\varphi_2);\ y = r_1 r_2\sin(\varphi_1+\varphi_2),$$

whence:

$$x = r_1 r_2(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) =$$
$$= r_1\cos\varphi_1\cdot r_2\cos\varphi_2 - r_1\sin\varphi_1\cdot r_2\sin\varphi_2 = a_1 a_2 - b_1 b_2,$$
$$y = r_1 r_2(\sin\varphi_1\cos\varphi_2 + \cos\varphi_1\sin\varphi_2) =$$
$$= r_1\sin\varphi_1\cdot r_2\cos\varphi_2 + r_1\cos\varphi_1\cdot r_2\sin\varphi_2 = b_1 a_2 + a_1 b_2,$$

so that we finally get:

$$(a_1 + b_1 i)(a_2 + b_2 i) = (a_1 a_2 - b_1 b_2) + (b_1 a_2 + a_1 b_2)\, i. \tag{7}$$

When $b_1 = b_2 = 0$, the factors are the real numbers a_1 and a_2 and the product reduces to the product of these numbers.

When $a_1 = a_2 = 0$, and $b_1 = b_2 = 1$, equation (7) gives:

$$i\cdot i = i^2 = -1,$$

i.e. *the square of imaginary unity is equal to* (-1).

Evaluation of positive integral powers of i thus gives:

$$i^2 = -1,\ i^3 = -i,\ i^4 = 1,\ i^5 = i,\ i^6 = -1,\ \ldots,$$

and in general, for any positive integer k:

$$i^{4k} = 1,\ i^{4k+1} = i,\ i^{4k+2} = -1,\ i^{4k+3} = -i.$$

The multiplication rule expressed by (7) can be stated as: *complex numbers are to be multiplied algebraically, taking* $i^2 = -1$.

If α is the complex number $a + bi$, the complex number $a - bi$ is called the conjugate of α, and is denoted by $\bar{\alpha}$.

In accordance with (3):

$$|\alpha|^2 = a^2 + b^2.$$

But from (7):

$$(a + bi)(a - bi) = a^2 + b^2,$$

so that:

$$|\alpha|^2 = (a + bi)(a - bi) = \alpha\bar{\alpha},$$

i.e. *the product of conjugate complex numbers is equal to the square of the modulus of each.*

We also note the obvious expressions:

$$\alpha + \bar{\alpha} = 2a; \; \alpha - \bar{\alpha} = 2bi. \tag{8}$$

It immediately follows from (4) and (7) that addition and multiplication of complex numbers obey the commutative law, i.e. a sum is independent of the order of its terms, and a product is independent of the order of its factors. It is easy to verify the associative and distributive laws, expressed by the identities:

$$(\alpha_1 + \alpha_2) + \alpha_3 = \alpha_1 + (\alpha_2 + \alpha_3); \; (\alpha_1 \alpha_2)\alpha_3 = \alpha_1(\alpha_2 \alpha_3);$$

$$(\alpha_1 + \alpha_2)\beta = \alpha_1\beta + \alpha_2\beta.$$

We leave the proof to the reader.

We remark finally that *the product of several factors has a modulus equal to the product of the moduli of the factors, and an amplitude equal to the sum of the amplitudes of the factors. Also, a product is zero when and only when at least one factor is zero.*

173. Division of complex numbers. Division is defined for complex numbers as the inverse operation to multiplication. It can easily be seen to follow that, if a complex number of modulus and amplitude (r_1, φ_1) is divided by a number of modulus and amplitude (r_2, φ_2), division yields a unique result if the divisor differs from zero, the modulus and amplitude of the quotient being r_1/r_2 and $(\varphi_1 - \varphi_2)$. We can write:

$$\frac{r_1(\cos\varphi_1 + i\sin\varphi_1)}{r_2(\cos\varphi_2 + i\sin\varphi_2)} = \frac{r_1}{r_2}[\cos(\varphi_1 - \varphi_2) + i\sin(\varphi_1 - \varphi_2)]. \tag{9}$$

Thus, *the modulus of the quotient of two complex numbers is equal to the quotient of their moduli, whilst the amplitude of the quotient is equal to the difference of their amplitudes.* If $r_2 = 0$, (9) becomes meaningless.

If the numbers are given as $a_1 + b_1 i$ and $a_2 + b_2 i$ instead of in the trigonometric form, their moduli and amplitudes can be expressed in terms of a_1, a_2, b_1, b_2 and set in (9), and the following expression obtained for the quotient:

$$\frac{a_1 + b_1 i}{a_2 + b_2 i} = \frac{a_1 a_2 + b_1 b_2}{a_2^2 + b_2^2} + \frac{b_1 a_2 - a_1 b_2}{a_2^2 + b_2^2} i,$$

This expression can also be obtained directly by considering i as irrational and multiplying numerator and denominator by the conjugate complex of the denominator, thus getting rid of irrationals in the denominator:

$$\frac{a_1 + b_1 i}{a_2 + b_2 i} = \frac{(a_1 + b_1 i)(a_2 - b_2 i)}{a_2^2 + b_2^2} = \frac{(a_1 a_2 + b_1 b_2) + (b_1 a_2 - a_1 b_2) i}{a_2^2 + b_2^2},$$

and finally:

$$\frac{a_1 + b_1 i}{a_2 + b_2 i} = \frac{a_1 a_2 + b_1 b_2}{a_2^2 + b_2^2} + \frac{b_1 a_2 - a_1 b_2}{a_2^2 + b_2^2} i. \tag{10}$$

We mentioned above [172] that the commutative, associative and distributive laws continue to be valid in the case of addition and multiplication of complex numbers; and it follows that all the well-known transformations in regard to real numbers that follow from these laws, also hold for expressions containing complex numbers. We include here: the rules for enclosing in and removal of brackets, elementary formulae, Newton's binomial formula with positive integral exponent, formulae relating to progressions, etc.

We note a further important property of expressions containing complex numbers, connected by the signs of the four primary operations. The following proposition is obtained at once from (4), (5), (7) and (10): *if all the numbers in a sum, product, difference or quotient are replaced by their conjugates, the result of the operations is also replaced by its conjugate.*

For example, on substituting $(-b_1)$ and $(-b_2)$ for b_1 and b_2 in (7), we get:

$$(a_1 - b_1 i)(a_2 - b_2 i) = (a_1 a_2 - b_1 b_2) - (b_1 a_2 + a_1 b_2) i.$$

This property will evidently hold for any sort of expression that contains complex numbers, connected by the signs of the four primary operations.

174. Raising to a power. We obtain the rule for raising a complex number to a positive integral power by applying (6) in the case of n equal factors:

$$[r(\cos\varphi + i\sin\varphi)]^n = r^n(\cos n\varphi + i\sin n\varphi)\,, \tag{11}$$

i.e. *a complex number is raised to a positive integral power by raising its modulus to the same power and multiplying its amplitude by the exponent of the power.*

We obtain de Moivre's theorem on putting $r = 1$ in (11):

$$(\cos\varphi + i\sin\varphi)^n = \cos n\varphi + i\sin n\varphi. \tag{12}$$

Examples. **1.** If we expand the right-hand side of (12) by Newton's binomial formula and equate real and imaginary parts in accordance with (2), we get expressions for $\cos n\varphi$ and $\sin n\varphi$ in powers of $\cos\varphi$ and $\sin\varphi$†:

$$\begin{aligned}\cos n\varphi = \cos^n\varphi - \binom{n}{2}\cos^{n-2}\varphi\sin^2\varphi + \\ + \binom{n}{4}\cos^{n-4}\varphi\sin^4\varphi + \ldots + (-1)^k\binom{n}{2k}\cos^{n-2k}\varphi\sin^{2k}\varphi + \\ + \ldots + \begin{cases}(-1)^{\frac{n}{2}}\sin^n\varphi & (n\text{ even})\\ (-1)^{\frac{n-1}{2}} n\cos\varphi\sin^{n-1}\varphi & (n\text{ odd})\end{cases}\end{aligned} \tag{13}$$

$$\begin{aligned}\sin n\varphi = \binom{n}{1}\cos^{n-1}\varphi\sin\varphi - \binom{n}{3}\cos^{n-3}\varphi\sin^3\varphi + \binom{n}{5}\cos^{n-5}\varphi\sin^5\varphi - \ldots + \\ + \ldots + (-1)^k\binom{n}{2k+1}\cos^{n-2k-1}\sin^{2k+1}\varphi + \ldots + \\ + \begin{cases}(-1)^{\frac{n-2}{2}} n\cos\varphi\sin^{n-1}\varphi & (n\text{ even})\\ (-1)^{\frac{n-1}{2}}\sin^n\varphi & (n\text{ odd}).\end{cases}\end{aligned}$$

In the particular case of $n = 3$, (12) becomes, after removing the brackets:

$$\cos^3\varphi + 3i\cos^2\varphi\sin\varphi - 3\cos\varphi\sin^2\varphi - i\sin^3\varphi = \cos 3\varphi + i\sin 3\varphi\,,$$

whence

$$\cos 3\varphi = \cos^3\varphi - 3\cos\varphi\sin^2\varphi;\quad \sin 3\varphi = 3\cos^2\varphi\sin\varphi - \sin^3\varphi.$$

† The symbol $\binom{n}{m}$ denotes the number of combinations of n elements in groups of m, i.e.

$$\binom{n}{m} = \frac{n(n-1)\ldots(n-m+1)}{1\cdot 2\ldots m} = \frac{n!}{m!\,(n-m)!}\,,$$

2. To sum the expressions:

$$A_n = 1 + r\cos\varphi + r^2\cos 2\varphi + \ldots + r^{n-1}\cos(n-1)\varphi\,,$$

$$B_n = r\sin\varphi + r^2\sin 2\varphi + \ldots + r^{n-1}\sin(n-1)\varphi\,.$$

We put:

$$z = r(\cos\varphi + i\sin\varphi)$$

and form the complex number:

$$A_n + B_n i = 1 + r(\cos\varphi + i\sin\varphi) +$$
$$+ r^2(\cos 2\varphi + i\sin 2\varphi) + \ldots + r^{n-1}[\cos(n-1)\varphi + i\sin(n-1)\varphi].$$

We use equation (11) and the formula for the sum of a geometrical progression:

$$A_n + B_n i = 1 + z + z^2 + \ldots + z^{n-1} = \frac{1-z^n}{1-z} = \frac{1-r^n(\cos\varphi + i\sin\varphi)^n}{1-r\ (\cos\varphi + i\sin\varphi)} =$$
$$= \frac{(1-r^n\cos n\varphi) - ir^n\sin n\varphi}{(1-r\cos\varphi) - ir\sin\varphi}\,.$$

On multiplying the numerator and denominator of the last fraction by $(1 - r\cos\varphi) + ir\sin\varphi$, the conjugate of the denominator, we get:

$$A_n + B_n i = \frac{[(1-r^n\cos n\varphi) - ir^n\sin n\varphi\,]\,[(1-r\cos\varphi) + ir\sin\varphi]}{(1-r\cos\varphi)^2 + r^2\sin^2\varphi} =$$
$$= \frac{(1-r^n\cos n\varphi)(1-r\cos\varphi) + r^{n+1}\sin\varphi\sin n\varphi}{r^2 - 2r\cos\varphi + 1} +$$
$$+ \frac{(1-r^n\cos n\varphi)\,r\sin\varphi - (1-r\cos\varphi)\,r^n\sin n\varphi}{r^2 - 2r\cos\varphi + 1}\,i =$$
$$= \frac{r^{n+1}\cos(n-1)\varphi - r^n\cos n\varphi - r\cos\varphi + 1}{r^2 - 2r\cos\varphi + 1} +$$
$$+ \frac{r^{n+1}\sin(n-1)\varphi - r^n\sin n\varphi + r\sin\varphi}{r^2 - 2r\cos\varphi + 1}\,i\,.$$

Equating real and imaginary parts in accordance with (2) gives:

$$A_n = 1 + r\cos\varphi + r^2\cos 2\varphi + \ldots + r^{n-1}\cos(n-1)\varphi =$$
$$= \frac{r^{n+1}\cos(n-1)\varphi - r^n\cos n\varphi - r\cos\varphi + 1}{r^2 - 2r\cos\varphi + 1}\,;$$

$$B_n = r\sin\varphi + r^2\sin 2\varphi + \ldots + r^{n-1}\sin(n-1)\varphi =$$
$$= \frac{r^{n+1}\sin(n-1)\varphi - r^n\sin n\varphi + r\sin\varphi}{r^2 - 2r\cos\varphi + 1}\,.$$

On taking the absolute value of the real number r as less than unity, and letting n increase indefinitely, we obtain in the limit the sum of the infinite series:

$$\left.\begin{aligned} 1 + r\cos\varphi + r^2\cos 2\varphi + \dots &= \frac{1 - r\cos\varphi}{r^2 - 2r\cos\varphi + 1}, \\ r\sin\varphi + r^2\sin 2\varphi + \dots &= \frac{r\sin\varphi}{r^2 - 2r\cos\varphi + 1}. \end{aligned}\right\} \tag{14}$$

If we set $r = 1$ in A_n and B_n, we get:

$$1 + \cos\varphi + \cos 2\varphi + \dots + \cos(n-1)\varphi = \frac{\cos(n-1)\varphi - \cos n\varphi - \cos\varphi + 1}{2(1-\cos\varphi)} =$$

$$= \frac{2\sin\frac{\varphi}{2}\left(n - \frac{1}{2}\right)\varphi + 2\sin^2\frac{\varphi}{2}}{4\sin^2\frac{\varphi}{2}} = \frac{\sin\left(n - \frac{1}{2}\right)\varphi + \sin\frac{\varphi}{2}}{2\sin\frac{\varphi}{2}} =$$

$$= \frac{\sin\frac{n\varphi}{2}\cos\frac{(n-1)\varphi}{2}}{\sin\frac{\varphi}{2}}. \tag{15_1}$$

We find similarly:

$$\sin\varphi + \sin 2\varphi + \dots + \sin(n-1)\varphi = \frac{\sin\frac{n\varphi}{2}\sin\frac{(n-1)\varphi}{2}}{\sin\frac{\varphi}{2}}. \tag{15_2}$$

175. Extraction of roots. *The n-th root of a complex number is defined as the complex number whose n-th power is the original number.*

The equation

$$\sqrt[n]{r(\cos\varphi + i\sin\varphi)} = \varrho(\cos\psi + i\sin\psi)$$

is thus equivalent to:

$$\varrho^n(\cos n\psi + i\sin n\psi) = r(\cos\varphi + i\sin\varphi).$$

But the moduli of equal complex numbers must be equal, and their amplitudes can only differ by a multiple of 2π, i.e.

$$\varrho^n = r,\ n\psi = \varphi + 2k\pi,$$

whence

$$\varrho = \sqrt[n]{r},\ \psi = \frac{\varphi + 2k\pi}{n},$$

where $\sqrt[n]{r}$ is the arithmetic value of the root, and k is any integer.

We thus have:

$$\sqrt[n]{r(\cos\varphi + i\sin\varphi)} = \sqrt[n]{r}\left(\cos\frac{\varphi + 2k\pi}{n} + i\sin\frac{\varphi + 2k\pi}{n}\right), \quad (16)$$

i.e. *the root of a complex number is to be extracted by extracting the root of its modulus and dividing its amplitude by the exponent of the root.*

All possible positive integral values can be taken by k in (16); it can be shown, however, that there are only n different values of the root, corresponding to:

$$k = 0, 1, 2, \ldots, (n-1). \quad (17)$$

We prove this by noting that the right-hand side of (16) has different values for two different values of k, $k = k_1$ and $k = k_2$, when the amplitudes $\frac{\varphi + 2k_1\pi}{n}$ and $\frac{\varphi + 2k_2\pi}{n}$ do not differ by a multiple of 2π, whilst they are the same if the amplitudes do differ by a multiple of 2π.

But the absolute value of the difference of two numbers $(k_1 - k_2)$ in the series (17) is less than n, and therefore the difference

$$\frac{\varphi + 2k_1\pi}{n} - \frac{\varphi + 2k_2\pi}{n} = \frac{k_1 - k_2}{n}\,2\pi$$

cannot be a multiple of 2π, i.e. the n different values of k of series (17) correspond to the n different values of the root. Now let k_2 be an integer not included in series (17). We can divide by n and write it in the form:

$$k_2 = qn + k_1,$$

where q is an integer and k_1 is one of the numbers of (17), and hence

$$\frac{\varphi + 2k_2\pi}{n} = \frac{\varphi + 2k_1\pi}{n} + 2\pi q$$

i.e. the same value of the root corresponds to k_2 as to k_1, belonging to (17). Thus, *the n-th root of a complex number has n different values.*

The only exception to this rule is the case of the number under the root being zero, i.e. $r = 0$, when all the above mentioned values of the root are zero.

Examples. **1.** We find all the values of $\sqrt[3]{i}$. The modulus of i is unity and its amplitude is $\pi/2$, and hence:

$$\sqrt[3]{i} = \sqrt[3]{\cos\frac{\pi}{2} + i\sin\frac{\pi}{2}} = \cos\frac{\frac{\pi}{2}+2k\pi}{3} + i\sin\frac{\frac{\pi}{2}+2k\pi}{3} \quad (k=0,1,2).$$

We get the following three values for $\sqrt[3]{i}$:

$$\cos\frac{\pi}{6} + i\sin\frac{\pi}{6} = \frac{\sqrt{3}}{2} + \frac{1}{2}i\,; \quad \cos\frac{5\pi}{6} + i\sin\frac{5\pi}{6} = -\frac{\sqrt{3}}{2} + \frac{1}{2}i\,;$$

$$\cos\frac{3\pi}{2} + i\sin\frac{3\pi}{2} = -i.$$

2. We consider all the values of $\sqrt[n]{1}$, i.e. all the solutions of the equation

$$z^n = 1.$$

The modulus of unity is unity, and its amplitude is zero, so that

$$\sqrt[n]{1} = \sqrt[n]{\cos 0 + i\sin 0} = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n} \quad (k=0,\ 1, 2, \ldots, n-1).$$

Let ε be the value obtained for the root with $k = 1$:

$$\varepsilon = \cos\frac{2\pi}{n} + i\sin\frac{2\pi}{n}.$$

By de Moivre's theorem:

$$\varepsilon^k = \cos\frac{2k\pi}{n} + i\sin\frac{2k\pi}{n},$$

i.e. all the roots of the equation $z^n = 1$ have the form

$$\varepsilon^k \ (k = 0, 1, 2, \ldots, n-1),$$

where we have to take $\varepsilon^0 = 1$.

We now consider an equation of the form

$$z^n = a,$$

We introduce a new variable u in place of z, putting

$$z = u\sqrt[n]{a},$$

where $\sqrt[n]{a}$ is one of the values of the nth root of a.

If we substitute this expression for z in the given equation, we get the equation for u:

$$u^n = 1.$$

It is clear from this that all the roots of $z^n = a$ can be put in the form:

$$\sqrt[n]{a}\,\varepsilon^k \qquad (k = 0, 1, 2, \ldots, n-1),$$

where $\sqrt[n]{a}$ is one of the n values of this root and ε^k takes all the values of the nth root of unity.

176. Exponential functions. We considered above the exponential function e^x in the case of a real exponent x. We now generalize the concept of exponential function to include a complex exponent. With a real exponent, e^x can be put in the form of a series [129]:

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots$$

We define the exponential function with a pure imaginary exponent by means of an analogous series, i.e. *we write :*

$$e^{yi} = 1 + \frac{yi}{1!} + \frac{(yi)^2}{2!} + \frac{(yi)^3}{3!} + \ldots$$

Separating the real and imaginary terms gives:

$$e^{yi} = \left(1 - \frac{y^2}{2!} + \frac{y^4}{4!} - \frac{y^6}{6!} + \ldots\right) + i\left(\frac{y}{1!} - \frac{y^3}{3!} + \frac{y^5}{5!} - \frac{y^7}{7!} + \ldots\right),$$

from which we deduce, on recalling the expansions of $\cos y$ and $\sin y$ [130]:

$$e^{yi} = \cos y + i \sin y. \tag{18}$$

This formula also defines an exponential function with pure imaginary exponent.

Substitution of $(-y)$ for y gives:

$$e^{-yi} = \cos y - i \sin y. \tag{19}$$

Solution of (18) and (19) with respect to $\cos y$ and $\sin y$ gives us Euler's formulae, expressing the trigonometric functions in terms of exponential functions with pure imaginary exponents:

$$\cos y = \frac{e^{yi} + e^{-yi}}{2}; \quad \sin y = \frac{e^{yi} - e^{-yi}}{2i}. \tag{20}$$

We obtain from (18) a new *exponential form of a complex number*, of modulus r and amplitude φ:

$$r(\cos\varphi + i\sin\varphi) = re^{\varphi i}.$$

The exponential function with a complex exponent $x + yi$ is defined by:

$$e^{x+yi} = e^x \cdot e^{yi} = e^x (\cos y + i \sin y), \tag{21}$$

i.e. *the modulus of* e^{x+yi} *is taken as* e^x, *and its amplitude as* y.

The rule for adding exponents when multiplying is easily generalized to include complex numbers.

Let $z = x + iy$ and $z_1 = x_1 + y_1 i$:

$$e^z \cdot e^{z_1} = e^x(\cos y + i \sin y) \cdot e^{x_1}(\cos y_1 + i \sin y_1),$$

or, using the rule for multiplying complex numbers [172]:

$$e^z \cdot e^{z_1} = e^{x+x_1}[\cos(y + y_1) + i \sin(y + y_1)].$$

But it follows from the definition of (21) that the expression on the right-hand side of this equation represents:

$$e^{(x+x_1)+(y+y_1)i}, \text{ i.e. } e^{z+z_1}.$$

The rule for subtracting exponents on division,

$$\frac{e^z}{e^{z_1}} = e^{z-z_1}$$

can be verified at once by multiplying the quotient by the denominator.

We have, for a positive integer n:

$$(e^z)^n = e^z e^z \dots e^z = e^{nz}.$$

We can use Euler's formula to express any positive integral power of $\sin \varphi$ and $\cos \varphi$, or any product of such powers, as the sum of terms containing only the first powers of the sines or cosines of multiples of the angle:

$$\sin^m \varphi = \frac{(e^{\varphi i} - e^{-\varphi i})^m}{2^m i^m}; \quad \cos^m \varphi = \frac{(e^{\varphi i} + e^{-\varphi i})^m}{2^m}. \tag{22}$$

We obtain the required expression on expanding the right-hand sides of these equations by Newton's binomial formula, cross-multiplying, and replacing the exponential by the trigonometric functions in the expansions obtained.

Examples. **1.**

$$\cos^4 \varphi = \frac{(e^{\varphi i} + e^{-\varphi i})^4}{16} = \frac{e^{4\varphi i}}{16} + \frac{4e^{2\varphi i}}{16} + \frac{6}{16} + \frac{4e^{-2\varphi i}}{16} + \frac{e^{-4\varphi i}}{16} =$$

$$= \frac{1}{8}\,\frac{e^{4\varphi i} + e^{-4\varphi i}}{2} + \frac{1}{2}\,\frac{e^{2\varphi i} + e^{-2\varphi i}}{2} + \frac{3}{8} = \frac{3}{8} + \frac{1}{2}\cos 2\varphi + \frac{1}{8}\cos 4\varphi.$$

2.

$$\sin^4\varphi\cos^3\varphi = \frac{(e^{\varphi i}-e^{-\varphi i})^4}{16}\cdot\frac{(e^{\varphi i}+e^{-\varphi i})^3}{8} = \frac{(e^{2\varphi i}-e^{-2\varphi i})^3(e^{\varphi i}-e^{-\varphi i})}{128} =$$

$$= \frac{(e^{6\varphi i}-3e^{2\varphi i}+3e^{-2\varphi i}-e^{-6\varphi i})(e^{\varphi i}-e^{-\varphi i})}{128} =$$

$$= \frac{e^{7\varphi i}-e^{5\varphi i}+3e^{3\varphi i}-3e^{\varphi i}+3e^{-\varphi i}-3e^{-3\varphi i}-e^{-5\varphi i}+e^{-7\varphi i}}{128} =$$

$$= \frac{3}{64}\cos\varphi - \frac{3}{64}\cos 3\varphi - \frac{1}{64}\cos 5\varphi - \frac{1}{64}\cos 7\varphi.$$

We remark here that any integral power of $\cos\varphi$ and any even power of $\sin\varphi$ represents an even function of φ, i.e. one whose value remains unchanged when φ is replaced by $(-\varphi)$, and which contains only cosines of multiples of the angle. An odd function of φ, i.e. one which changes sign when φ is replaced by $(-\varphi)$, is obtained, for example, in the case of odd powers of $\sin\varphi$; its expansion contains only sines of multiples of the angle, and the absolute term is entirely missing. All these details will be considered in more detail in our treatment of trigonometric series.

177. Trigonometric and hyperbolic functions. We have so far considered trigonometric functions only in the case of real arguments. We define these functions for any complex argument z by means of Euler's formula:

$$\cos z = \frac{e^{zi}+e^{-zi}}{2}; \quad \sin z = \frac{e^{zi}-e^{-zi}}{2i},$$

where the expressions on the right have the meanings indicated in [176].

By using these formulae, together with the basic properties of exponential functions, the trigonometric formulae are easily shown to hold for a complex argument. We propose the proof of the following relationships as an exercise to the reader:

$$\sin^2 z + \cos^2 z = 1;$$

$$\sin(z+z_1) = \sin z\cos z_1 + \cos z\sin z_1;$$

$$\cos(z+z_1) = \cos z\cos z_1 - \sin z\sin z_1;$$

The functions $\tan z$ and $\cot z$ are defined by:

$$\tan z = \frac{\sin z}{\cos z} = \frac{1}{i}\cdot\frac{e^{zi}-e^{-zi}}{e^{zi}+e^{-zi}} = \frac{1}{i}\cdot\frac{e^{2zi}-1}{e^{2zi}+1};$$

$$\cot z = \frac{\cos z}{\sin z} = = i\,\frac{e^{zi}+e^{-zi}}{e^{zi}-e^{-zi}} = i\frac{e^{2zi}+1}{e^{2zi}-1}.$$

We now introduce the *hyperbolic functions*.

Hyperbolic sin and cos are defined by:

$$\sinh z = \frac{\sin iz}{i} = \frac{e^z - e^{-z}}{2}, \quad \cosh z = \cos iz = \frac{e^z + e^{-z}}{2};$$

$$\tanh z = \frac{\sinh z}{\cosh z} = \frac{e^z - e^{-z}}{e^z + e^{-z}} = \frac{e^{2z} - 1}{e^{2z} + 1};$$

$$\coth z = \frac{\cosh z}{\sinh z} = \frac{e^z + e^{-z}}{e^z - e^{-z}} = \frac{e^{2z} + 1}{e^{2z} - 1}.$$

The following relationships are easily shown to hold by using these formulae:

$$\left.\begin{aligned}
&\cosh^2 z - \sinh^2 z = 1\,;\\
&\sinh(z_1 \pm z_2) = \sinh z_1 \cosh z_2 \pm \cosh z_1 \sinh z_2;\\
&\cosh(z_1 \pm z_2) = \cosh z_2 \cosh z_2 \pm \sinh z_1 \sinh z_2\,;\\
&\sinh 2z = 2 \sinh z \cosh z, \cosh 2z = \cosh^2 z + \sinh^2 z;\\
&\tanh 2z = \frac{2 \tanh z}{1 + \tanh^2 z}, \quad \coth 2z = \frac{1 + \coth^2 z}{2 \coth z}.
\end{aligned}\right\} \tag{23}$$

The above relationships are analogous to the relationships in the ordinary trigonometry of angles, and give rise to *hyperbolic trigonometry*. The formulae of this latter are obtained from the corresponding formulae of ordinary trigonometry by replacing $\sin z$ by $i \sinh z$, and $\cos z$ by $\cosh z$; this follows at once from the formulae defining the hyperbolic functions.

By using the above principle, we easily obtain the following formulae for reducing the sums of hyperbolic functions to the logarithmic form:

$$\left.\begin{aligned}
&\sinh z_1 + \sinh z_2 = 2 \sinh \frac{z_1 + z_2}{2} \cosh \frac{z_1 - z_2}{2}\,;\\
&\sinh z_1 - \sinh z_2 = 2 \sinh \frac{z_1 - z_2}{2} \cosh \frac{z_1 + z_2}{2}\,;\\
&\cosh z_1 + \cosh z_2 = 2 \cosh \frac{z_1 + z_2}{2} \cosh \frac{z_1 - z_2}{2}\,;\\
&\cosh z_1 - \cosh z_2 = 2 \sinh \frac{z_1 + z_2}{2} \sinh \frac{z_1 - z_2}{2}.
\end{aligned}\right\} \tag{24}$$

We now consider the hyperbolic functions with real values of the argument:

$$\sinh x = \frac{e^x - e^{-x}}{2}; \quad \cosh x = \frac{e^x + e^{-x}}{2},$$

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}; \quad \coth x = \frac{e^{2x} + 1}{e^{2x} - 1}.$$

The graph of $y = \cosh x$ is a catenary [78], which is dealt with in greater detail in [178].

The graphs of $\cosh x$, $\sinh x$, $\tanh x$ and $\coth x$ are illustrated in Fig. 171.

Direct differentiation gives us the following expressions for the derivatives:

$$\frac{d \sinh x}{dx} = \cosh x\,; \quad \frac{d \cosh x}{dx} = \sinh x\,;$$

$$\frac{d \tanh x}{dx} = \frac{1}{\cosh^2 x}\,; \quad \frac{d \coth x}{dx} = -\frac{1}{\sinh^2 x}.$$

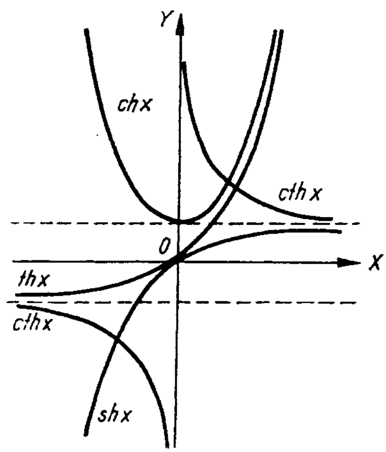

FIG. 171

Hence we obtain the table of integrals:

$$\int \sinh x\,dx = \cosh x + C; \quad \int \cosh x\,dx = \sinh x + C;$$

$$\int \frac{dx}{\cosh^2 x} = \tanh x + C\,; \quad \int \frac{dx}{\sinh^2 x} = -\coth x + C.$$

The name "hyperbolic functions" is due to the fact that $\cosh t$ and $\sinh t$ play the same role in the parametric form of the *rectangular hyperbola*

$$x^2 - y^2 = a^2,$$

as $\cos t$ and $\sin t$ play as regards the circle

$$x^2 + y^2 = a^2.$$

The parametric form of the circle is

$$x = a\cos t;\ y = a\sin t\,,$$

whilst that of the rectangular hyperbola is

$$x = a\cosh t;\ y = a\sinh t\,,$$

as is easily seen on using the relationship:

$$\cosh^2 t - \sinh^2 t = 1\,.$$

The geometrical significance of the parameter t is similar in both these cases. If S denotes the area of the sector AOM (Fig. 172), and

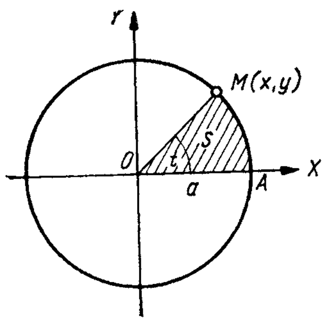

FIG. 172

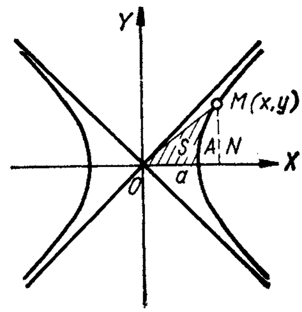

FIG. 173

S_0 denotes the area of the circle ($S_0 = \pi a^2$), we evidently have

$$t = 2\pi\frac{S}{S_0}\cdot$$

Let S now denote the area of the corresponding sector of the rectangular hyperbola (Fig. 173). We have:

$$S = \text{area of } OMN - \text{area of } AMN = \frac{1}{2}xy - \int_a^x y\,dx =$$

$$= \frac{1}{2}x\sqrt{x^2 - a^2} - \int_a^x \sqrt{x^2 - a^2}\,dx\,.$$

On evaluating the integral in accordance with the formula of [92], we find:

$$S = \frac{1}{2}x\sqrt{x^2 - a^2} - \frac{1}{2}\left[x\sqrt{x^2 - a^2} - a^2\log\left(x + \sqrt{x^2 - a^2}\right)\right]_a^x =$$

$$= \frac{1}{2}a^2\log\left(\frac{x}{a} + \sqrt{\frac{x^2}{a^2} - 1}\right).$$

If we now put

$$t = 2\pi \frac{S}{S_0} = \log\left(\frac{x}{a} + \sqrt{\frac{x^2}{a^2} - 1}\right),$$

where S_0 again denotes the area of the circle, we easily find:

$$e^t = \frac{x}{a} + \sqrt{\frac{x^2}{a^2} - 1},$$

$$e^{-t} = \frac{1}{\frac{x}{a} + \sqrt{\frac{x^2}{a^2} - 1}} = \frac{x}{a} - \sqrt{\frac{x^2}{a^2} - 1},$$

and on adding and multiplying the result by $a/2$;

$$x = \frac{a}{2}(e^t + e^{-t}) = a \cosh t,$$

$$y = \sqrt{x^2 - a^2} = \sqrt{a^2 \cosh^2 t - a^2} = a \sinh t,$$

i.e. we obtain the parametric form of the rectangular hyperbola.

178. The catenary. We investigate the curve in which a flexible cord of uniform density hangs when supported at its ends A_1 and A_2 (Fig. 174).

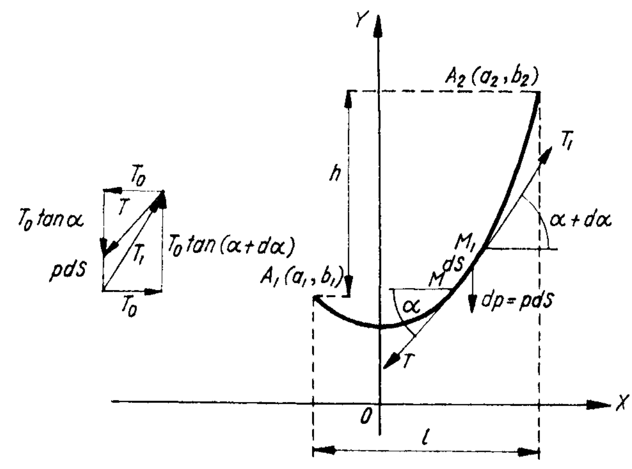

FIG. 174

We take the axis OX along the horizontal in the plane of the curve, and OY vertically upwards. We divide the cord into elements $MM_1 = ds$. The forces acting on each element are the tensions T and T_1 from the remainder of the cord and the weight of the element. The tensions act along the tangents at the ends M and M_1 of the element (T is in the negative direction

of the tangent, T_1 in the positive direction). The weight of an element is proportional to its length:

$$dp = \varrho ds ,$$

where ϱ is the linear density of the cord (the weight of unit length.)

The necessary and sufficient conditions for equilibrium are that the sums of the horizontal and vertical components of the forces acting on the element are both zero. The horizontal component of the weight dp of the element is zero, so that the horizontal components of T and T_1 must be equal in magnitude and opposite in sign. Let T_0 denote the common magnitude of these horizontal components.

The following expressions for the vertical components of the tensions are obtained from the figure:

$$-T_0 \tan a = -T_0 y' \text{ and } T_0 \tan (a + da) = T_0(y' + dy').$$

Here, da denotes the increment of the angle a, formed by the tangen- with OX, when we pass from the point M to M_1, and dy' is the correspondt ing increment in the slope of the tangent, i.e. in $\tan a$.

We find, on equating to zero the components of T, T_1 and the weight ϱds along OY:

$$T_0(y' + dy') - T_0 y' - \varrho ds = 0 ,$$

i.e.

$$T_0 dy' = \varrho ds ,$$

which can be written as:

$$T_0 dy' = \varrho \sqrt{1 + y'^2}\, dx. \tag{25}$$

We now separate the variables [93]:

$$\frac{dy'}{\sqrt{1 + y'^2}} = \frac{dx}{k}, \text{ where } k = \frac{T_0}{\varrho};$$

it is to be noted that k is a constant, directly proportional to the horizontal component of the tension, and inversely proportional to the linear density of the cord.

We integrate the equation obtained:

$$\log (y' + \sqrt{1 + y'^2}) = \frac{x + C_1}{k} ,$$

whence

$$e^{\frac{x+C_1}{k}} = y' + \sqrt{1 + y'^2};$$

then we find y' by introducing the reciprocal:

$$e^{-\frac{x+C_1}{k}} = \frac{1}{y' + \sqrt{1 + y'^2}} = \sqrt{1 + y'^2} - y'.$$

We subtract this equation from the previous one; which gives

$$y' = \frac{1}{2}\left(e^{\frac{x+C_1}{k}} - e^{-\frac{x+C_1}{k}}\right).$$

On integrating a second time, we get the required equation for the curve of the cord:

$$y + C_2 = \frac{k}{2}\left(e^{\frac{x+C_1}{k}} + e^{-\frac{x+C_1}{k}}\right). \tag{26}$$

The arbitrary constants C_1 and C_2 are found from the fact that the curve passes through the points $A_1(a_1, b_1)$ and $A_2(a_2, b_2)$. In practice, however, the main interest is not so much in the actual equation of the curve, i.e. in the constants C_1, C_2, as in the relationship between the horizontal and vertical distances between the points of suspension and the length of arc $A_1 A_2$.

When studying the relationship between these three magnitudes, we can of course make a parallel displacement of the coordinate axes. We can put the origin at the point $(-C_1, -C_2)$, in which case, from (26), $C_1 = C_2 = 0$, and (26) can be written more simply:

$$y = \frac{k}{2}\left(e^{\frac{x}{k}} + e^{-\frac{x}{k}}\right) = k \cosh \frac{x}{k}, \tag{26_1}$$

from which it is clear that *the equation of the suspension is a catenary.*

Let the coordinates of A_1 and A_2, with the above choice of axes, be (a_1, b_1) and (a_2, b_2). Let l, h, s denote respectively the horizontal and vertical distances between the points of suspension, and the length of the cord; then

$$l = a_2 - a_1; \quad h = b_2 - b_1 = k\left(\cosh \frac{a_2}{k} - \cosh \frac{a_1}{k}\right),$$

$$s = \int_{a_1}^{a_2} \sqrt{1 + y'^2}\, dx = \int_{a_1}^{a_2} \sqrt{1 + \sinh^2 \frac{x}{k}}\, dx = \int_{a_1}^{a_2} \cosh \frac{x}{h}\, dx = k\left(\sinh \frac{a_2}{k} - \sinh \frac{a_1}{k}\right).$$

We find on using (24):

$$h = 2k \sinh \frac{a_2 + a_2}{2k} \sinh \frac{a_2 - a_1}{2k} = 2k \sinh \frac{l}{2k} \sinh \frac{a_2 + a_1}{2k},$$

$$s = 2k \sinh \frac{a_2 - a_1}{2k} \cosh \frac{a_2 + a_1}{2k} = 2k \sinh \frac{l}{2k} \cosh \frac{a_2 + a_1}{2k},$$

which gives us, on using the first of relationships (23):

$$s^2 - h^2 = 4k^2 \sinh^2 \frac{l}{2k},$$

representing the required relationship between l, h and s.

This can also be written in the following form:

$$\frac{\sinh \frac{l}{2k}}{\frac{l}{2k}} = \frac{\sqrt{s^2 - h^2}}{l}. \tag{27}$$

If the points of suspension and the length of the cord are given, the magnitudes l, h and s are known, and we obtain an equation for the para-

meter k; if the linear density ϱ of the cord is also known, (27) can be used to find the horizontal component of the tension T_0.

For the sake of brevity, we put:

$$\frac{l}{2k} = \xi; \qquad \frac{\sqrt{s^2 - h^2}}{l} = c.$$

Equation (27) becomes:

$$\frac{\sinh \xi}{\xi} = c. \tag{27_1}$$

The exponential function can be expanded into a power series [129], so that we can write

$$\frac{\sinh \xi}{\xi} = \frac{e^{\xi} - e^{-\xi}}{2\xi} = 1 + \frac{\xi^2}{3!} + \frac{\xi^4}{5!} + \frac{\xi^6}{7!} + \ldots,$$

from which it is clear that, as ξ increases from 0 to $+\infty$, the ratio also increases steadily, from 1 to $+\infty$. Accordingly, for every given $c > 1$, (27_1) has one positive root, which can be obtained from tables of hyperbolic functions.†

The given magnitudes l, h, s must here satisfy the condition:

$$c = \frac{\sqrt{s^2 - h^2}}{l} \geqslant 1 \text{ or } s^2 \geqslant h^2 + l^2,$$

which is also obvious geometrically, since $\sqrt{h^2 + l^2}$ is the length of the chord $A_1 A_2$, whilst s is the length of the catenary between these points.

For example, let:

$s = 100$ m., $l = 50$ m., $h = 20$ m., $\varrho = 20$ kg/m.

We get:

$$c = 0.02\sqrt{10{,}000 - 400} = 0.8\sqrt{6} = 1.96$$

and we find the root of (27_1) from the tables:

$$\xi = \frac{l}{2k} = 2.15,$$

whence

$$T^0 = k\varrho = \frac{k}{2\xi}\varrho = \frac{50}{2 \times 2.15} \cdot 20 = 232 \text{ kg}.$$

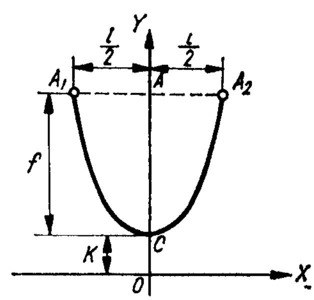

FIG. 175

Let the points of suspension be at the same height. We investigate the sag f of the cord (Fig. 175):

$$f = \overline{OA} - \overline{OC} = \frac{k}{2}\left(e^{\frac{l}{2k}} + e^{-\frac{l}{2k}}\right) - k = \frac{k}{2}\left(e^{\frac{l}{2k}} + e^{-\frac{l}{2k}} - 2\right).$$

† The tables of Jahnke and Emde, for instance.

Expansion of the exponential functions gives us:

$$f = \frac{1}{2!}\frac{l^2}{2^2 \cdot k} + \frac{1}{4!}\frac{l^4}{2^4 \cdot k^3} + \cdots \tag{28}$$

We have in exactly the same way for $s = \text{arc}\, A_1 A_2$ [with $h = 0$ in (27)]:

$$s = 2k \sinh \frac{l}{2k} = k\left(e^{\frac{l}{2k}} - e^{-\frac{l}{2k}}\right) = l + \frac{1}{3!}\frac{l^3}{2^2 \cdot k^2} + \frac{1}{5!}\frac{l^5}{2^4 \cdot k^4} + \cdots \tag{29}$$

We can find k approximately by confining ourselves to the first term of series (28):

$$k \sim \frac{l^2}{8f}.$$

We preserve the first two terms of expansion (29) and substitute the expression found for k:

$$s \sim l + \frac{8}{3}\frac{f^2}{l}.$$

Differentiation of this expression gives us the *relationship between the increase in length of the cord and the increase in sag :*

$$ds \sim \frac{16}{3}\frac{f\,df}{l}, \quad \text{or} \quad df \sim \frac{3l}{16f}\,ds.$$

We obtained equation (25) on the assumption that the gravity force acting on each element of the cord was proportional to the length of the element. In certain cases, e.g. when considering the chains of suspension bridges, the gravity force must be reckoned proportional to the length of the projection of the element on the horizontal axis. This happens when the loading due to the roadway of the bridge is large compared with the weight of the chains themselves, so that the latter can be neglected. We have now, instead of equation (25):

$$T_0\, dy' = \varrho dx,$$

whence

$$y' = \frac{\varrho}{T_0}x + C_1,$$

and

$$y = \frac{\varrho}{2T_0}x^2 + C_1 x + C_2,$$

i.e. *the curve of the suspension is a parabola.*

If we suppose that the ends A_1 and A_2 of the chain (or cord) are at the same level, whilst we locate the origin of coordinates at the vertex of the parabola (Fig. 176), its equation will be:

$$y = ax^2 \quad \left(a = \frac{\varrho}{2T_0}\right).$$

We find, as above, the length of the span $l = \overline{A_1 A_2}$ and the sag $f = OA$. We get from the equation of the parabola:

$$f = a\frac{l^2}{4},$$

whence

$$a = \frac{4f}{l^2}.$$

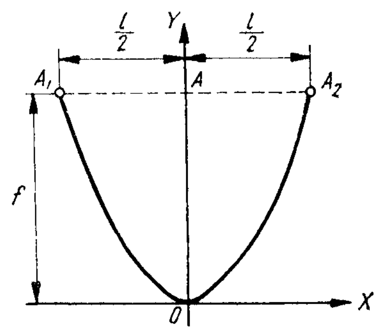

FIG. 176

We calculate the length of arc A_1A_2, which is twice the length of arc OA_2:

$$s = 2\int_0^{l/2} \sqrt{1 + 4a^2x^2}\, dx.$$

We have by Newton's binomial formula:

$$\sqrt{1 + 4a^2 x^2} = (1 + 4a^2 x^2)^{1/2} = 1 + 2a^2 x^2 - 2a^4 x^4 + \ldots$$

and integration gives us the expansion for s:

$$s = l + \frac{1}{6}a^2l^3 - \frac{1}{40}a^4l^5 + \ldots$$

We substitute the expression found above for a:

$$s = l + \frac{8}{3}\left(\frac{f}{l}\right)^2 l - \frac{32}{5}\left(\frac{f}{l}\right)^4 l + \ldots = l\left[1 + \frac{8}{3}\varepsilon^2 - \frac{32}{5}\varepsilon^4 + \ldots\right],$$

where $\varepsilon = f/l$. If we confine ourselves to the first two terms of this series, we obtain the approximate formula:

$$s \sim l + \frac{8}{3}\frac{f^2}{l},$$

which is the same as the corresponding formula for the catenary.

179. Logarithms. *The natural logarithm of the complex number $r(\cos\varphi + i\sin\varphi)$ is defined as the power to which* e *must be raised to give the number.* If we use Log to denote natural logarithms, we can say that the equation

$$\operatorname{Log}[r(\cos\varphi + i\sin\varphi)] = x + yi$$

is equivalent to:

$$e^{x+yi} = r(\cos\varphi + i\sin\varphi).$$

The latter equation can be written:

$$e^x(\cos y + i\sin y) = r(\cos\varphi + i\sin\varphi),$$

whence we find, on equating moduli and amplitudes:

$$e^x = r, y = \varphi + 2k\pi \ (k = 0, \pm 1, \pm 2, \ldots) ,$$

i.e.

$$x = \log r \text{ and } x + yi = \log r + (\varphi + 2k\pi)\, i$$

and finally:

$$\mathrm{Log}[r(\cos\varphi + i\sin\varphi)] = \log r + (\varphi + 2k\pi)\, i , \tag{30}$$

i.e. *the natural logarithm of a complex number is the complex number whose real part is the ordinary logarithm of the modulus and whose imaginary part is the product of i and one of the values of the amplitude.*

It follows that the natural logarithm of any number has an infinite set of values. The only exception is zero, the logarithm of which does not exist. If we make the amplitude obey the inequality

$$-\pi < \varphi \leqslant \pi ,$$

we obtain the so-called *principal values of the logarithm*. The principal value of the logarithm is denoted by log instead of Log, to distinguish it from the general value, as given by (30); thus

$$\log[r(\cos\varphi + i\sin\varphi)] = \log r + \varphi\, i , \tag{31}$$

where

$$-\pi < \varphi \leqslant \pi .$$

We define *the complex power of a complex number* with the aid of logarithms. If u and v are two complex numbers, where $u \neq 0$, we put:

$$u^v = e^{v\,\mathrm{Log}\,u} .$$

We note that Log u, and therefore u^v, has in general an infinite set of values.

Examples. **1.** The modulus of i is unity and its amplitude is $\frac{1}{2}\pi$, so that

$$\mathrm{Log}\, i = \left(\frac{1}{2}\pi + 2k\pi\right) i \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

2. We find i^i:

$$i^i = e^{i \log i} = e^{-\left(\frac{1}{2}\pi + 2k\pi\right)} \qquad (k = 0, \pm 1, \pm 2, \ldots).$$

180. Sinusoidal quantities and vector diagrams. The application of complex quantities to the study of harmonic oscillations may be mentioned. We consider a variable current j, whose value at each instant is the same throughout the circuit, and is given by:

$$j = j_m \sin(\omega t + \varphi) , \tag{32}$$

where t represents time, and j_m, ω and φ are constants.

The constant j_m, reckoned positive, is called the *amplitude*, ω is the *frequency*, connected with the *period* T by the formula

$$T = \frac{2\pi}{\omega},$$

and φ is the *phase* of the alternating current.

A current whose instantaneous value varies in accordance with (32) is said to be *sinusoidal*. What has been said also applies to a voltage:

$$v = v_m \sin(\omega t + \varphi_1), \tag{33}$$

and our discussion below is concerned with sinusoidal currents and voltages, as given by (32) and (33).

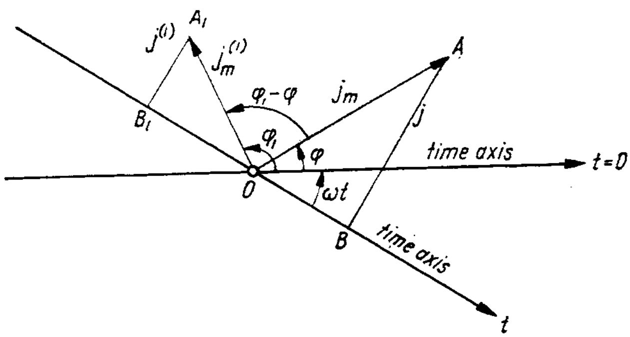

FIG. 177

There is a simple method of representing geometrically sinusoidal quantities of the same frequency. We take a radius vector through a fixed point O of the plane, and let it rotate clockwise with angular velocity ω; this radius vector is referred to as the *time axis*.

Let the time axis coincide with axis OX initially, i.e. at $t = 0$.

We draw a vector $\overline{OA}$ (Fig. 177) of length j_m, forming an angle φ with the initial position of the time axis (recalling that angles are reckoned positive in the counter-clockwise direction). At time t, $\overline{OA}$ will form an angle $(\varphi + \omega t)$ with the time axis, which has rotated through an angle ωt. We clearly get the quantity $j = j_m \sin(\omega t + \varphi)$ from the projection of $\overline{OA}$ on the perpendicular to the time axis, obtained by turning this latter counter-clockwise through an angle $\pi/2$; or more briefly, j is obtained by dropping a perpendicular from the end of OA on to the time axis, and taking its length with the appropriate sign.

Another sinusoidal quantity of the same period,

$$j^{(1)} = j_m^{(1)} \sin(\omega t + \varphi_1),$$

must be represented by drawing a vector of length $j_m^{(1)}$ at an angle to the first equal to

$$\psi = \varphi_1 - \varphi.$$

Sinusoidal quantities of the same frequency can thus be represented by fixed vectors in a plane. The length of each vector gives the corresponding amplitude, whilst the angle between any two represents their phase difference. This type of construction is called a *vector diagram* of the system of sinusoidal quantities of the same period.

The sum of several such *sinusoidal quantities* is also sinusoidal and of the same period, being represented by the sum of the corresponding vectors on the diagram.

The definition of multiplication given in [172] can be used to put operations with vector diagrams into a convenient analytic form.

We shall in future denote vectors by letters in heavy type.

We shall reckon the product of a vector $\mathbf{j}$ *and a complex number* $re^{\varphi i}$ *as the vector obtained by multiplying the length of* $\mathbf{j}$ *by* r *and turning its direction through an angle* φ, i.e. we multiply the complex number representing $\mathbf{j}$ by the complex number $re^{\varphi i}$ in accordance with the rule given in [172].

If $re^{\varphi i}$ is given in the form $(a + bi)$, we can write the product as the sum of two vectors:

$$(a + bi)\,\mathbf{j} = a\mathbf{j} + bi\mathbf{j}\,,$$

the first term being a vector parallel to $\mathbf{j}$, and the second a vector perpendicular to $\mathbf{j}$.

We can split any vector $\mathbf{j}_1$ into two mutually perpendicular components, and write it in the form:

$$\mathbf{j}_1 = a\mathbf{j} + bi\mathbf{j} = (a + bi)\,\mathbf{j}\,.$$

Here, $|\,a + bi\,|$ evidently gives the ratio of $\mathbf{j}$ to $\mathbf{j}_1$, whilst the amplitude of $(a + bi)$ represents the angle between the two vectors. This angle gives the phase difference of the quantities corresponding to the vectors.

We bring in the concept of the *mean square* value, denoted by $M(j^2)$, of the sinusoidal quantity (32). This is defined by the equation:

$$M\,(j^2) = \frac{1}{T}\int\limits_0^T j^2\,\mathrm{d}t\,.$$

We have:

$$j^2 = j_m^2 \sin^2(\omega t + \varphi) = \frac{1}{2}\,j_m^2 - \frac{1}{2}\,j_m^2 \cos 2\,(\omega t + \varphi)$$

and integration from 0 to $T = 2\pi/\omega$ gives:

$$M\,(j^2) = \frac{1}{2}\,j_m^2 - \left[\frac{1}{4\omega}\,j_m^2 \sin 2\,(\omega t + \varphi)\right]_0^{2\pi/\omega} = \frac{1}{2}\,j_m^2\,.$$

The square root of the mean square value is called the *effective*, or *root mean square*, value of the quantity:

$$j_{\mathrm{eff}} = \sqrt{M\,(j^2)} = \frac{j_m}{\sqrt{2}}\,.$$

In practice, each vector in a diagram is usually made equal in length to the effective value of the quantity, and not to its amplitude, i.e. the lengths

of the vectors are diminished in the ratio 1: $\sqrt{2}$ by comparison with the construction given above.

We obtain by differentiating (32):

$$\frac{\mathrm{d}j}{\mathrm{d}t} = \omega j_m \cos(\omega t + \varphi) = \omega j_m \sin\left(\omega t + \varphi + \frac{\pi}{2}\right),$$

i.e. the derivative $\mathrm{d}\mathbf{j}/\mathrm{d}t$ only differs from $\mathbf{j}$ in that its amplitude is multiplied by ω, and its phase increased by $\pi/2$.

The relationship deduced becomes in vector notation:

$$\frac{\mathrm{d}\mathbf{j}}{\mathrm{d}t} = \omega i \mathbf{j}. \tag{34}$$

Integration of (32) gives us:

$$\int j\,\mathrm{d}t = -\frac{1}{\omega} j_m \cos(\omega t + \varphi) = \frac{1}{\omega} j_m \sin\left(\omega t + \varphi - \frac{1}{2}\pi\right),$$

where the arbitrary constant has to be neglected if we want to obtain a quantity that is also sinusoidal; and it follows from this that:†

$$\int \mathbf{j}\,\mathrm{d}t = \frac{1}{\omega i}\mathbf{j}. \tag{35}$$

181. Examples. 1. We consider an alternating current circuit consisting of a resistance R in series with an inductance L and a capacity C. If v denotes voltage and j denotes current, we know from physics that

$$v = Rj + L\frac{\mathrm{d}j}{\mathrm{d}t} + \frac{1}{C}\int j\,\mathrm{d}t.$$

We confine ourselves for the present to the *steady state* conditions, and to the case when the voltage and current are sinusoidal quantities of the same period. The above equation can be written in vector form, with v and j replaced by the vectors of voltage and current, $\mathbf{v}$ and $\mathbf{j}$:

$$\mathbf{v} = R\mathbf{j} + L\frac{\mathrm{d}\mathbf{j}}{\mathrm{d}t} + \frac{1}{C}\int \mathbf{j}\,\mathrm{d}t;$$

We find from this on recalling (34) and (35):

$$\mathbf{v} = R\mathbf{j} + \omega L i\mathbf{j} + \frac{1}{\omega C i}\mathbf{j} = (R + ui)\,\mathbf{j} = \zeta\mathbf{j}, \tag{36}$$

where

$$u = \omega L - \frac{1}{\omega C}; \text{ and } \zeta = R + ui. \tag{37}$$

The relationship obtained between the voltage and current vectors has the ordinary form of Ohm's law, except for the fact that the ohmic resistance is here replaced by a complex factor, called the *impedance of the*

† The symbol $\mathrm{d}\mathbf{j}/\mathrm{d}t$ denotes the vector corresponding to the sinusoidal quantity $\mathrm{d}j/\mathrm{d}t$, whilst $\int \mathbf{j}\mathrm{d}t$ denotes the vector corresponding to $\int j\mathrm{d}t$.

circuit, and made up of three "impedances": an *ohmic resistance* R, *an inductive impedance* (ωLi) and *a capacitative impedance* $(1/\omega Ci)$.

We also get from (36) the resolution of vector $\mathbf{v}$ into two components: $R\mathbf{j}$ in the direction of $\mathbf{j}$, and $ui\mathbf{j}$ perpendicular to $\mathbf{j}$. The first is called the *real*, and the second is called the *wattless*, component of the voltage. These terms become clear if we calculate the *mean power* W in our circuit, which we define as the arithmetic mean over a full period of the instantaneous power vj:

$$W = \frac{1}{T}\int_0^T vj\,\mathrm{d}t = \frac{v_m j_m}{T}\int_0^T \sin(\omega t + \varphi_1)\sin(\omega t + \varphi_2)\,\mathrm{d}t\,;$$

here, φ_1 denotes the phase of the voltage, and φ_2 the phase of the current, so that

$$v = v_m \sin(\omega t + \varphi_1)\,; \quad j = j_m \sin(\omega t + \varphi_2)\,.$$

We easily find:

$$W = \frac{v_m j_m}{2T}\int_0^T [\cos(\varphi_1 - \varphi_2) - \cos(2\omega t + \varphi_1 + \varphi_2)]\,\mathrm{d}t$$

$$= \frac{v_m j_m}{2}\cos(\varphi_1 - \varphi_2) = v_{eff}\, j_{eff}\cos(\varphi_1 - \varphi_2)\,. \tag{38}$$

The greatest absolute value of mean power is thus obtained when the phases of the voltage and current either coincide or differ by π, whilst the least, zero, power is found when the phases differ by $\pi/2$.

The wattless component $ui\mathbf{j}$ of the vector $\mathbf{v}$ gives a mean power of zero when put into this expression for W, since $ui\mathbf{j}$ is perpendicular to $\mathbf{j}$, i.e. $\cos(\varphi_1 - \varphi_2) = 0$ in this case; thus, all the mean power that passes into Joule heat is provided by the real component alone.

We can write (36) as:

$$\mathbf{j} = \frac{1}{\zeta}\mathbf{v} = \eta\mathbf{v}\,, \text{ where } \eta = \frac{1}{R + ui} = g + hi\,,$$

or

$$\mathbf{j} = g\mathbf{v} + hi\mathbf{v}\,.$$

The complex factor η is called the *admittance* of the circuit, and is *equal to the reciprocal of the impedance* This last formula expresses the current vector as the sum of a real component in the direction of $\mathbf{v}$, and a wattless component perpendicular to $\mathbf{v}$.

2. The basic rules, derived from Ohm's and Kirchhoff's laws, for finding the resistance of a direct current circuit that includes several resistances in series or parallel, can also be used for steady sinusoidal alternating current circuits, provided we agree to replace the instantaneous values of voltage and current by the corresponding vectors, and the ohmic resistances by the impedances.

Thus, if the circuit includes the impedances in series:

$$\zeta_1 = R_1 + x_1 i;\ \zeta_2 = R_2 + x_2 i;\ \ldots,$$

the voltage and current vectors will be connected by:

$$\mathbf{v} = \zeta' \mathbf{j}, \text{ where } \zeta' = \zeta_1 + \zeta_2 + \dots, \tag{39}$$

i.e. *impedances are added in a series circuit.*

If the same impedances are in parallel, we have on the other hand:

$$\mathbf{v} = \zeta'' \mathbf{j}, \quad \text{where } \frac{1}{\zeta''} = \frac{1}{\zeta_1} + \frac{1}{\zeta_2} + \dots, \tag{40}$$

i.e. *the admittances are added in a parallel circuit.*

The total impedances in the case of series connection of impedances $\zeta_1, \zeta_2, \dots,$ is found geometrically by the simple process of constructing the geometrical sum of the vectors representing these complex numbers.

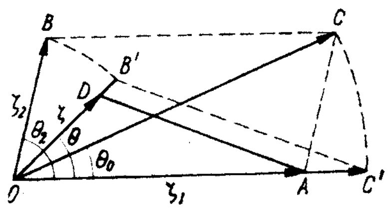

FIG. 178

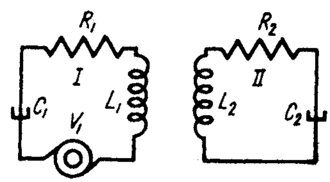

FIG. 179

We indicate the construction in the case of two apparent resistances ζ_1 and ζ_2, connected in parallel. We have by the above rule:

$$\zeta'' = \frac{1}{\dfrac{1}{\zeta_1} + \dfrac{1}{\zeta_2}} = \frac{\zeta_1 \zeta_2}{\zeta_1 + \zeta_2}.$$

If we put:

$$\zeta'' = \varrho e^{\theta i}, \quad \zeta_1 = \varrho_1 e^{\theta_1 i}; \quad \zeta_2 = \varrho_2 e^{\theta_2 i}; \quad \zeta_1 + \zeta_2 = \varrho_0 e^{\theta_0 i},$$

we have:

$$\varrho = \frac{\varrho_1 \varrho_2}{\varrho_0}; \quad \theta = \theta_1 + \theta_2 - \theta_0.$$

This leads us to the following geometrical construction (Fig. 178).†

We first find the sum $\zeta_1 + \zeta_2 = \overline{OC}$; then we draw ΔAOD, similar to ΔCOB, by turning ΔCOB to the position $C'OB'$ and drawing $\overline{AD} \| \overline{C'B'}$.

We deduce from the fact that the triangles are similar that:

$$\overline{OD} = \overline{OA}\,\frac{\overline{OB}}{\overline{OC}}, \quad \text{i.e.} \quad \varrho = \frac{\varrho_1 \varrho_2}{\varrho_0}, \quad \theta = \theta_2 - \theta_0 \ \ (\theta_1 = 0),$$

which is what we wanted to show.

3. We consider the coupled oscillations in two magnetically coupled circuits (Fig. 179). Let v_1, j_1 denote the external electromotive force and

† We have simplified the figure by taking the axis OX along the vector ζ_1, amounting to the assumption that $\theta_1 = 0$. In the general case, it is sufficient to turn OX clockwise through an angle θ_1.

the current in circuit I, and j_2 the current in circuit II (with no external electro-motive force); let R_1, R_2, L_1, L_2, C_1, C_2 be respectively the resistances, inductances and capacities in the circuits, and M the mutual inductance between the circuits.

We have the relationships:

$$v_1 = R_1 j_1 + L_1 \frac{\mathrm{d}j_1}{\mathrm{d}t} + M \frac{\mathrm{d}j_2}{\mathrm{d}t} + \frac{1}{C_1} \int j_1 \,\mathrm{d}t,$$

$$0 = R_2 j_2 + L_2 \frac{\mathrm{d}j_2}{\mathrm{d}t} + M \frac{\mathrm{d}j_1}{\mathrm{d}t} + \frac{1}{C_2} \int j_2 \,\mathrm{d}t.$$

When we consider the steady state, in which the currents and voltages vary sinusoidally at the same frequency, these equations can be written in vector form:

$$\mathbf{v}_1 = \left(R_1 + \omega L_1 i + \frac{1}{\omega C_1 i}\right) \mathbf{j}_1 + \omega M i \mathbf{j}_2 = \zeta_1 \mathbf{j}_1 + \omega M i \mathbf{j}_2,$$

$$0 = \omega M i \mathbf{j}_1 + \left(R_2 + \omega L_2 i + \frac{1}{\omega C_2 i}\right) \mathbf{j}_2 = \omega M i \mathbf{j}_1 + \zeta_2,$$

where ζ_1 and ζ_2 are the impedances of circuits I and II when taken separately.

We can easily solve for $\mathbf{j}_1$ and $\mathbf{j}_2$:

$$\mathbf{j}_1 = \frac{\zeta_2}{\zeta_1\zeta_2 + \omega^2 M^2} \mathbf{v}_1; \quad \mathbf{j}_2 = -\frac{\omega M i}{\zeta_1\zeta_2 + \omega^2 M^2} \mathbf{v}_1.$$

On re-writing the first equation as:

$$\mathbf{v}_1 = \left(\zeta_1 + \frac{\omega^2 M^2}{\zeta_2}\right) \mathbf{j}_1,$$

we can say that the presence of circuit II changes the impedance of circuit I by an amount $\omega^2 M^2/\zeta_2$.

182. Curves in the complex form. If we agree to represent real numbers by points on a given axis OX, the variation of a real variable is expressed by a corresponding displacement of the point on OX. In exactly the same way, the variation of a complex variable $\zeta = x + yi$ amounts to displacement of the corresponding point on the *plane* XOY.

The case of particular interest is when the point ζ describes a certain curve during its variation; this occurs when the real and imaginary parts, i.e. the coordinates x and y, are functions of some parameter u, which we shall take to be real:

$$x = \varphi_1(u); \; y = \varphi_2(u). \tag{41}$$

We shall now write simply:

$$\zeta = f(u), \text{ where } f(u) = \varphi_1(u) + i\varphi_2(u),$$

and this equation will be referred to as *the equation of curve* (41) *in the complex form*.

Equations (41) give the parametric form of the curve in *rectangular* coordinates. We arrive at the form in *polar* coordinates by writing ζ in the exponential form:

$$\zeta = \varrho e^{\theta i}; \ \varrho = \psi_1(u), \ \theta = \psi_2(u).$$

The factor ϱ in this expression is the same as $|\zeta|$, whilst $e^{\theta i}$, which coincides with the "sign" (± 1) in the case of real $\zeta (\theta = 0$ or $\pi)$, is a vector of unit length, denoted by the symbol:

$$\operatorname{sgn} \zeta = e^{\theta i} = \frac{\zeta}{|\zeta|}$$

(sgn is an abbreviation of the Latin word "signum" — sign.)

The necessity for considering the equations of curves in the complex form arises out of vector diagrams. If we take the current vector $\mathbf{j}$ as constant in the relationship

$$\mathbf{v} = \zeta \mathbf{j},$$

but allow one of the circuit constants to change, the impedance ζ and the vector $\mathbf{v}$ will also change; the end of vector $\mathbf{v}$ describes a curve, called the *voltage diagram*, and the construction of this gives us a clear picture of the variation of $\mathbf{v}$. The point ζ also describes a curve (the *impendance diagram)*, which differs from the voltage diagram only in the choice of scale (taking the vector $\mathbf{j}$ as unity).

We now consider the equations of some simple curves.

1. The equation of a straight line, passing through the point $\zeta_0 = x_0 + y_0 i$, and forming an angle a with OX, is:

$$\zeta = \zeta_0 + u e^{ai};$$

the parameter u here denotes the distance from point ζ_0 to ζ.

2. The equation of a circle with its centre at ζ_0 and radius r is:

$$\zeta = \zeta_0 + r e^{ui}.$$

3. An ellipse with its centre at the origin, its major axis along OX, and semi-axes a and b, has the equation in the complex form [177]:

$$\zeta = x + yi = a \cos u + bi \sin u = \frac{1}{2}(a + b) e^{ui} + \frac{1}{2}(a-b) e^{-ui}.$$

If the major axis makes an angle φ_0 with OX, the equation of the ellipse takes the form:

$$\zeta = e^{\varphi_0 i}\left[\frac{1}{2}(a + b) e^{ui} + \frac{1}{2}(a-b) e^{-ui}\right].$$

In the general case, when the centre of the ellipse is at ζ_0 and its major axis makes an angle φ_0 with OX, its equation is:

$$\zeta = \zeta_0 + e^{\varphi_0 i}\left[\frac{1}{2}(a + b) e^{ui} + \frac{1}{2}(a - b) e^{-ui}\right].$$

If $b = a$, this equation becomes the equation of a circle of radius a:

$$\zeta = \zeta_0 + a e^{(\varphi_0 + u) i},$$

where $(\varphi_0 + u)$ is a real parameter like u.

If $b = 0$, we get a segment of a straight line:

$$\zeta = \zeta_0 + a e^{\varphi_0 i} \frac{1}{2} (e^{ui} + e^{-ui}) = \zeta_0 + a e^{\varphi_0 i} \cos u \; ; \quad \zeta = \zeta_0 + v e^{\varphi_0 i},$$

making an angle φ_0 with OX, of length $2a$, and with its centre at ζ_0, since the parameter $v = a \cos u$ is real, like u, but can only take values between $(-a)$ and $(+a)$.

If we consider the circle and the straight line segment as limiting cases of an ellipse, when its minor semi-axis becomes equal to the major, or becomes zero, we can say in general that *the equation*

$$\zeta = \zeta_0 + \mu_1 e^{ui} + \mu_2 e^{-ui}, \tag{42}$$

where ζ_0, μ_1 and μ_2 *are any desired complex numbers, always represents the equation of an ellipse.*

This follows by putting:

$$\mu_1 = M_1 e^{\theta_1 i} \; ; \quad \mu_2 = M_2 e^{\theta_2 i} \; ; \; \frac{1}{2} (\theta_1 + \theta_2) = \varphi_0 \; ; \quad \frac{1}{2} (\theta_1 - \theta_2) = \theta_0,$$

when (42) can be written in the form:

$$\zeta = \zeta_0 + M_1 e^{(u+\theta_1) i} + M_2 e^{-(u-\theta_2) i} = \zeta_0 + e^{\varphi_0 i} [M_1 e^{(u+\theta_0) i} + M_2 e^{-(u+\theta_0) i}],$$

from which it is clear that the curve in question is in fact an ellipse with its centre at ζ_0, semi-axes $(M_1 \pm M_2)$, and its major axis making an angle φ_0 with OX, i.e. in a direction bisecting the angle between the vectors μ_1 and μ_2. The ellipse becomes a circle for $M_2 = 0$, and a segment of a line for $M_2 = M_1$.

4. Curves of great value in the study of alternating current phenomena in circuits with continuously distributed of resistance, capacity and inductance, are those with equations of the complex form:

$$\zeta = \nu e^{\gamma u}, \tag{43}$$

where ν and γ are any desired complex constants.

We have here, on putting $\nu = N_1 e^{\varphi_0 i}$, $\gamma = a + bi$ and using polar coordinates:

$$\zeta = \varrho e^{\theta i} = N_1 e^{\varphi_0 i} e^{(a+bi) u} = N_1 e^{au} e^{(bu+\varphi_0) i},$$

that is:

$$\varrho = N_1 e^{au}, \quad \theta = bu + \varphi_0 ,$$

whence

$$u = \frac{\theta - \varphi_0}{b},$$

or finally:

$$\varrho = N e^{a\theta/b} \quad (N = N_1 e^{a\varphi_0/b}) ,$$

i.e. the curve in question is a logarithmic spiral (Fig. 180), for the case $a/b > 0$).

More difficult curves of the type:

$$\zeta = \nu_1 e^{\gamma_1 u} + \nu_2 e^{\gamma_2 u} + \ldots + \nu_s e^{\gamma_s u}$$

can be obtained by constructing the "component spirals":

$$\zeta_1 = \nu_1 e^{\gamma_1 u}, \ \zeta_2 = \nu_2 e^{\gamma_2 u}, \ldots, \ \zeta_s = \nu_s e^{\gamma_s u},$$

and working out geometrically, for each value of u, the sum of the corresponding values of $\zeta_1, \zeta_2, \ldots, \zeta_s$ (Fig. 181).

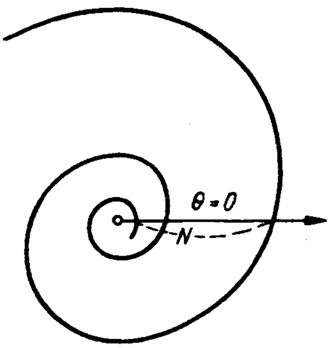

FIG. 180

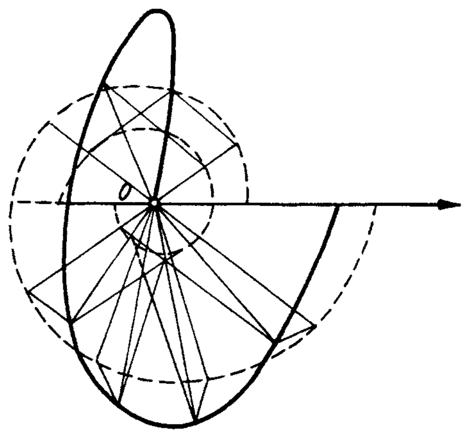

FIG. 181

183. Representation of harmonic oscillations in complex form. A damped harmonic oscillation is expressed by the formula:

$$x = A e^{-\varepsilon t} \sin (\omega t + \varphi_0) , \tag{44}$$

where A and ε are positive constants. We introduce the complex quantity here:

$$\zeta = A e^{\left(\varphi_0 - \frac{1}{2}\pi\right) i} e^{(\omega + \varepsilon i) it} = A e^{-\varepsilon t + \left(\omega t + \varphi_0 - \frac{1}{2}\pi\right) i} . \tag{45}$$

The real part of this complex quantity coincides with the expression (44). We can thus represent any damped harmonic oscillation as the real part of a complex expression of the form:

$$\zeta = \alpha e^{\beta i t},$$

where α and β are complex numbers. In the case of (45):

$$\alpha = A e^{\left(\varphi_0 - \frac{1}{2}\pi\right) i} \quad \text{and} \quad \beta = \omega + \varepsilon i .$$

For a pure harmonic oscillation without damping, $\varepsilon = 0$, and β becomes a real number.

Expression (45) for ζ coincides with (43) when

$$\nu = Ae^{\left(\varphi_0 - \frac{1}{2}\pi\right)i}, \quad \gamma = (\omega + \varepsilon i)\, i = -\varepsilon + \omega i \quad \text{and} \quad u = t.$$

It is clear from this that the point ζ describes a logarithmic spiral as t varies, the polar angle θ being a linear function of time t:

$$\theta = \omega t + \varphi_0 - \frac{1}{2}\pi,$$

i.e. the radius vector from the origin to ζ rotates about the origin with constant angular velocity ω. The projection of ζ on OX performs the damped oscillation (44). If $\varepsilon = 0$, ζ moves on the circle $\varrho = A$, and its projection on OX moves according to the law of harmonic oscillation without damping:

$$x = A \sin(\omega t + \varphi_0).$$

§ 18. Basic properties and evaluation of the zeros of integral polynomials

184. Algebraic equations. We undertake the study in the present article of the integral polynomial:

$$f(z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_k z^{n-k} + \ldots + a_{n-1} z + a_n,$$

where $a_0, a_1, \ldots, a^k, \ldots, a_n$ are given complex numbers, z is a complex variable, and the initial coefficient a_0 can be assumed to differ from zero. The primary operations with polynomials are familiar from elementary algebra. We shall only recall the basic result as regards division. If the degree of a polynomial $\varphi(z)$ is not higher than the degree of another polynomial $f(z)$, $f(z)$ can be expressed in the form:

$$f(z) = \varphi(z) \cdot Q(z) + R(z),$$

where $Q(z)$ and $R(z)$ are also polynomials, the degree of $R(z)$ being lower than that of $\varphi(z)$. We speak of $Q(z)$ and $R(z)$ respectively as the quotient and remainder of the division of $f(z)$ by $\varphi(z)$. The quotient and remainder are completely determinate polynomials, so that $f(z)$ is expressed uniquely in terms of $\varphi(z)$ in the above form.

The zeros of the polynomial are defined as the values of z which, when substituted in the polynomial, cause it to vanish. The zeros of $f(z)$ are thus the roots of the equation:

$$f(z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_k z^{n-k} + \ldots + a_{n-1} z + a_n = 0. \quad (1)$$

The equation written is called an *algebraic equation of the* nth *degree.*

If $f(z)$ is divided by $(z - a)$, the quotient $Q(z)$ is a polynomial of degree $(n - 1)$ with initial coefficient a_0, whilst the remainder R will not contain z. The identity follows from the basic definition of division:

$$f(z) = (z - a)\, Q(z) + R.$$

We get on substituting $z = a$ in this identity:

$$R = f(a)\ ,$$

i.e. *division of a polynomial* $f(z)$ *by* $(z - a)$ *gives a remainder equal to* $f(a)$ (Bézout's theorem).

In particular, a necessary and sufficient condition that $f(z)$ be divisible by $(z - a)$ with no remainder is that

$$f(a) = 0\ ,$$

i.e. *the necessary and sufficient condition for the polynomial to be divisible without remainder by the linear factor* $(z - a)$ *is that* $z = a$ *be a zero of the polynomial.*

Thus, knowing that $z = a$ is a zero of $f(z)$, we can divide the polynomial by the factor $(z - a)$:

$$f(z) = (z - a)\, f_1(z)\ ,$$

where

$$f_1(z) = b_0 z^{n-1} + b_1 z^{n-2} + \ldots + b_{n-2} z + b_{n-1} \quad (b_0 = a_0)\ ;$$

finding the remaining zeros leads to the solution of the equation:

$$b_0 z^{n-1} + b_1 z^{n-2} + \ldots + b_{n-2} z + b_{n-1} = 0$$

of degree $(n - 1)$.

Before proceeding, we have to answer the question: does every algebraic equation possess roots? The answer can be in the negative, as regards non-algebraic equations; for instance, the equation

$$e^z = 0 \quad (z = x + yi)$$

never has a root, since the modulus e^x of the left-hand side does not vanish for any x. In the case of algebraic equations, however, our question has an affirmative answer, contained in the following basic theorem of algebra: *every algebraic equation has at least one real or complex root.*

We state this theorem here without proof. The proof will be found in the third volume, on the theory of functions of a complex variable.

185. Factorization of polynomials. In accordance with the basic theorem, every polynomial

$$f(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n \,, \tag{2}$$

has a zero $z = z_1$; it can thus be divided by $(z - z_1)$ and written [184]:

$$f(z) = (z - z_1)(a_0 z^{n-1} + \dots).$$

The second factor on the right-hand side of this equation has a root $z = z_2$, by the same theorem, and can be divided by $(z - z_2)$; so that we can now write:

$$f(z) = (z - z_1)(z - z_2)(a_0 z^{n-2} + \dots).$$

On continuing to divide out the linear factors, we finally obtain the factorial form of $f(z)$:

$$f(z) = a_0(z - z_1)(z - z_2) \dots (z - z_n) \,, \tag{3}$$

i.e. *every polynomial of the n-th degree can be split into* $(n + 1)$ *factors, one of which is the initial coefficient, whilst the remainder are linear factors of the form* $(z - a)$.

At least one of the factors of (3) vanishes on substituting $z = z_s$ $(s = 1, 2, \dots, n)$, so that these $z = z_s$ *are zeros of* $f(z)$.

A z differing from all the z_s cannot be a zero of $f(z)$, since none of the factors of (3) vanishes for this z.

If all the z_s are different, $f(z)$ has precisely n different zeros; if some z_s are the same, the number of different zeros of $f(z)$ is less than n.

We can thus state the theorem: *a polynomial of degree n (or algebraic equation of degree n) cannot have more than n different zeros.*

An immediate consequence of this theorem is the proposition: *if a polynomial of degree not higher than n is known to have more than n different zeros, all its coefficients and its free term must be zero,* i.e. *the polynomial is identically zero.*

Suppose that two polynomials $f_1(z)$ and $f_2(z)$, of degrees not higher than n, have the same value for more than n different values of z. Their difference, $f_1(z) - f_2(z)$, is a polynomial of degree not higher than n, with more than n different roots, and therefore identically zero; $f_1(z)$ and $f_2(z)$ thus have the same coefficients. *If two polynomials,*

of degrees not higher than n, have the same values for more than n different values of z, they have the same coefficients and the same absolute term, i.e. *they are identical.*

This property of polynomials lies at the basis of the so-called method of undetermined coefficients, which we shall make use of later. The essence of the method consists in the fact that the identity of two polynomials implies that they have the same coefficients for the same powers of z.

We obtained the factor form (3) by dividing the polynomial $f(z)$ by linear factors in a definite order. We now show that the final form is independent of the way in which we divide out the factors, i.e. that *a polynomial has a unique factor form* (3).

Suppose that, in addition to (3), we have the factor form:

$$f(z) = b_0(z - z_1')(z - z_2') \dots (z - z_n'). \qquad (3_1)$$

On comparing these two forms, we can write the identity:

$$a_0(z - z_1)(z - z_2) \dots (z - z_n) = b_0(z - z_1')(z - z_2') \dots (z - z_n').$$

The left-hand side of this identity vanishes for $z = z_1$, and the same must therefore be true of the right-hand side, i.e. at least one of the z_k' must be equal to z_1. We can take, for exanple, $z_1 = z_1'$. We cancel $(z - z_1)$ from both sides and get the equation:

$$a_0(z - z_2) \dots (z - z_n) = b_0(z - z_2')(\dots (z - z_n').$$

This is valid for all z, including $z = z_1$. But this equation must now also be an identity, by the proposition proved above. We can use the same argument as above to show that $z_2' = z_2$, and so on, and finally, that $b_0 = a_0$, i.e. the form (3_1) must coincide with (3).

186. Multiple zeros. Some of the z_s appearing in (3) may be the same, as already mentioned. We can put the same factors together in groups, and write:

$$f(z) = a_0(z - z_1)^{k_1}(z - z_2)^{k_2} \dots (z - z_m)^{k_m}, \qquad (4)$$

where $z_1, z_2, \dots, z_m$ are different, and

$$k_1 + k_2 + \dots + k_m = n. \qquad (5)$$

If there is a factor $(z - z_s)^{k_s}$ in the factor form written, $z = z_s$ is called *a zero of order* k_s, and in general, *the zero* $z = a$ *of the*

polynomial $f(z)$ *is called a zero of order* k, *if* $f(z)$ *is divisible by* $(z-a)^k$, *but not by* $(z-a)^{k+1}$.

We now give another test for multiple (repeated) zeros. We introduce Taylor's theorem here, and first of all remark that the derivatives of the polynomial $f(z)$ can be defined by means of the same formulae as were valid in the case of a real variable:

$$f(z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_k z^{n-k} + \ldots + a_{n-1} z + a_n;$$

$$f'(z) = na_0 z^{n-1} + (n-1) a_1 z^{n-2} + \ldots + (n-k) a_k z^{n-k-1} + \\ + \ldots + a_{n-1};$$

$$f''(z) = n(n-1) a_0 z^{n-2} + (n-1)(n-2) a_1 z^{n-3} + \ldots + \\ + (n-k)(n-k-1) a_k z^{n-k-2} + \ldots + 2.1\, a_{n-2}.$$

Taylor's formula:

$$f(z) = f(a) + \frac{z-a}{1!} f'(a) + \frac{(z-a)^2}{2!} f''(a) + \\ + \ldots + \frac{(z-a)^k}{k!} f^{(k)}(a) + \ldots + \frac{(z-a)^n}{n!} f^{(n)}(a) \qquad (6)$$

consists of an elementary algebraic identity in a and z, valid for complex, as well as real, values of these symbols.

We now deduce the condition for $z = a$ to be a zero of order k of $f(z)$. We write (6) in the form:

$$f(z) = (z-a)^k \left[\frac{1}{k!} f^{(k)}(a) + \frac{z-a}{(k+1)!} f^{(k+1)}(a) + \right. \\ \left. + \ldots + \frac{(z-a)^{n-k}}{n!} f^{(n)}(a)\right] \\ + \left[f(a) + \frac{z-a}{1!} f'(a) + \ldots + \frac{(z-a)^{(k-1)}}{(k-1)!} f^{(k-1)}(a)\right].$$

The second polynomial in square brackets on the right is of lower degree than $(z-a)^k$, so that clearly [184], the first square bracket is the quotient, and the second the remainder, of the division of $f(z)$ by $(z-a)^k$. The necessary and sufficient condition for $f(z)$ to be divisible by $(z-a)^k$ is that this remainder should be identically zero. We consider it as a polynomial in $(z-a)$, and get the following conditions:

$$f(a) = f'(a) = \ldots = f^{(k-1)}(a) = 0. \qquad (7)$$

We must add to these the condition

$$f^{(k)}(a) \neq 0\,, \tag{8}$$

since if $f^{(k)}(a) = 0, f(z)$ is divisible by $(z - a)^{k+1}$ as well as by $(z - a)^k$. All in all, (7) *and* (8) *give the necessary and sufficient conditions for* $z = a$ *to be a zero of order* k *of* $f(z)$.

We put $\psi(z) = f'(z)$, so that:

$$\psi'(z) = f''(z);\ \psi''(z) = f'''(z);\ \dots;\ \psi^{(s-1)}(z) = f^{(s)}(z).$$

If $z = a$ is a zero of order k of $f(z)$, we have by (7) and (8):

$$\psi(a) = \psi'(a) = \dots = \psi^{(k-2)}(a) = 0 \text{ and } \psi^{(k-1)}(a) \neq 0\,,$$

i.e. $z = a$ is a zero of order $(k - 1)$ of $\psi(z)$ or, what amounts the same thing, of $f'(z)$, i.e. *a zero of order* k *of a given polynomial is a zero of order* $(k - 1)$ *of its first derivative*. It can be seen by using this property successively that *it is a zero of order* $(k - 2)$ *of the second derivative, a zero of order* $(k - 3)$ *of the third derivative, etc., and finally, a single, or simple, root of the* $(k - 1)$*-th derivative*.

Thus, if $f(z)$ has the factor form:

$$f(z) = a_0 (z - z_1)^{k_1} (z - z_2)^{k_2} \dots (z - z_m)^{k_m}, \tag{9}$$

$f'(z)$ will have the form:

$$f'(z) = (z - z_1)^{k_1-1} (z - z_2)^{k_2-1} \dots (z - z_m)^{k_m-1}\, \omega(z), \tag{10}$$

where $\omega(z)$ is a polynomial with no zeros in common with $f(z)$.

187. Horner's rule. We now give a practical rule for the evaluation of $f(z)$ and its derivatives for a given $z = a$.

Let the division of $f(z)$ by $(z - a)$ give a quotient $f_1(z)$ and a remainder r_1; let the division of $f_1(z)$ by $(z - a)$ give a quotient $f_2(z)$ and a remainder r_2; and so on:

$$\begin{aligned} f(z) &= (z - a) f_1(z) + r_1; \quad r_1 = f(a);\\ f_1(z) &= (z - a) f_2(z) + r_2; \quad r_2 = f_1(a);\\ f_2(z) &= (z - a) f_3(z) + r_3; \quad r_3 = f_2(a);\\ &\dots\dots\dots\dots\dots\dots\dots\dots \end{aligned}$$

We write (6) in the form:

$$f(z) = f(a) + (z - a)\left[\frac{f'(a)}{1} + \frac{f''(a)}{2!}(z - a) + \dots + \frac{f^{(n)}(a)}{n!}(z - a)^{n-1}\right].$$

We compare this formula with the first of the equations written above, and obtain:

$$f_1(z) = \frac{f'(a)}{1} + \frac{f''(a)}{2!}(z-a) + \ldots + \frac{f^{(n)}(a)}{n!}(z-a)^{n-1};\ r_1 = f(a).$$

We find, on treating $f_1(z)$ in the same way:

$$f_2(z) = \frac{f''(a)}{2!} + \frac{f'''(a)}{3!}(z-a) + \ldots + \frac{f^{(n)}(a)}{n!}(z-a)^{n-2};\ r_2 = \frac{f'(a)}{1}.$$

and in general:

$$r_{k+1} = \frac{f^{(n)}(a)}{k!} \quad (k = 1, 2, \ldots, n).$$

We now put:

$$f(z) = a_0 z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n;$$
$$f_1(z) = b_0 z^{n-1} + b_1 z^{n-2} + \ldots + b_{n-2} z + b_{n-1};\quad b_n = r_1$$

and show how the coefficients b_s of the quotient, and b_n of the remainder, can be evaluated. We remove the brackets and collect terms in the same power of z, and obtain the identity:

$$\begin{aligned} a_0 z^n + a_1 z^{n-1} + \ldots + a_{n-1} z + a_n = (z-a)(b_0 z^{n-1} + b_1 z^{n-2} + \ldots + \\ + b_{n-2} z + b_{n-1}) + b_n = b_0 z^n + (b_1 - b_0 a) z^{n-1} + (b_2 - b_1 a) z^{n-2} + \ldots + \\ + (b_{n-1} - b_{n-2} a) z + (b_n - b_{n-1} a), \end{aligned}$$

and comparison of the coefficients of like powers of z gives:

$$a_0 = b_0;\ a_1 = b_1 - b_0 a;\ a_2 = b_2 - b_1 a;\ \ldots;\ a_{n-1} = b_{n-1} - b_{n-2} a;$$
$$a_n = b_n - b_{n-1} a,$$

whence:

$$b_0 = a_0;\ b_1 = b_0 a + a_1;\ b_2 = b_1 a + a_2;\ \ldots;\ b_{n-1} = b_{n-2} a + a_{n-1};$$
$$b_n = b_{n-1} a + a_n = r_1.$$

These equations enable us successively to calculate the b_s.

Similarly, on writing the quotient and remainder of the division of $f_1(z)$ by $(z-a)$ as:

$$f_2(z) = c_0 z^{n-2} + c_1 z^{n-3} + \ldots + c_{n-3} z + c_{n-2};\quad c_{n-1} = r_2,$$

we obtain:

$$c_0 = b_0;\ c_1 = c_0 a + b_1;\ c_2 = c_1 a + b_2;\ \ldots;\ c_{n-2} = c_{n-3} a + b_{n-2};$$
$$c_{n-1} = c_{n-2} a + b_{n-1} = r_2,$$

i.e. coefficients c_s are calculated successively with the aid of the b_s, just as the b_s were calculated with the aid of the a_s.

This method of calculation is called Horner's rule.

We obtain the quantities $f^{(k)}(a)/k!$ on applying this rule.

The calculation is set out in the table below, which is self-explanatory.

a	$a_0,\quad a_1,\quad a_2,\quad a_3,\quad \ldots,\quad a_{n-2},\quad a_{n-1},\quad a_n$	
	$+\quad b_0a,\quad b_1a,\quad b_2a,\quad \ldots,\quad b_{n-3}a,\quad b_{n-2}a,\quad b_{n-1}a$	
	$b_0 = a_0,\quad b_1,\quad b_2,\quad b_3,\quad \ldots,\quad b_{n-2},\quad b_{n-1},$	$b_n = r_1 = f(a)$
	$+\quad c_0a,\quad c_1a,\quad c_2a,\quad \ldots,\quad c_{n-3}a,\quad c_{n-2}a$	
	$c_0 = a_0,\quad c_1,\quad c_2,\quad c_3,\quad \ldots,\quad c_{n-2},$	$c_{n-1} = r_2 = \dfrac{f'(a)}{1}$
. . . . . .	. . . .	. . . . . .
	$l_0 = a_0,\quad l_1$	$l_2 = r_{n-1} = \dfrac{f^{(n-2)}(a)}{(n-2)!}$
	$+\quad m_0a$	
	$m_0 = a_0$	$m_1 = r_n = \dfrac{f^{(n-1)}(a)}{(n-1)!}$
		$m_0 = \dfrac{f^{(n)}(a)}{n!}$

Example. To find the values of

$$f(z) = z^5 + 2z^4 - 2z^2 - 25z + 100$$

and its derivatives for $z = -5$.

$a = -5$	$1,\quad 2,\quad 0,\quad -2,\quad -25,\quad 100$	
	$-5,\quad 15,\quad -75,\quad 385,\quad -1800$	
	$1,\quad -3,\quad 15,\quad -77,\quad 360$	$-1700 = f(-5)$
	$-5,\quad 40,\quad -275,\quad 1760$	
	$1,\quad -8,\quad 55,\quad -352$	$2120 = \dfrac{f'(-5)}{1!}$
	$-5,\quad 65,\quad -600$	
	$1,\quad -13,\quad 120$	$-952 = \dfrac{f''(-5)}{2!}$
	$-5,\quad 90$	
	$1,\quad -18$	$210 = \dfrac{f'''(-5)}{3!}$
	-5	
	1	$-23 = \dfrac{f^{(\mathrm{iv})}(-5)}{4!}$
		$1 = \dfrac{f^{(\mathrm{v})}(-5)}{5!}$

188. Highest common factor. We take two polynomials $f_1(z)$ and $f_2(z)$, each of which must have a determinate factor of the form (3). The *highest common factor* of the two polynomials is defined as the product of all the linear factors of the type $(z - a)$ that appear in the expansions of both $f_1(z)$ and $f_2(z)$, each individual factor being raised to the lower of the powers to which it is raised in the expansions concerned. Constant factors have no part in the composition of the highest common factor. The highest common factor of two polynomials is a third polynomial, the zeros of which are in common with those of the first two, each zero having an order equal to the lower of the orders with which it appears in the two original polynomials. If the two original polynomials have no common zeros they are said to be *relatively prime*. The highest common factor of several polynomials can be defined in precisely the same way.

A knowledge of the expansion into linear factors of the given polynomials is essential, if the highest common factor is to be obtained by the above method. Yet obtaining the expansion (3) amounts to solving the equation $f(z) = 0$, and this represents one of the basic problems of algebra.

The highest common factor can be obtained by another method, however, similar to that used in arithmetic for finding the highest common factor of integers; this is the *method of successive division* [Euclid's algorithm], which does not require factorization. Let the degree of $f_1(z)$ be not lower than that of $f_2(z)$. We divide the first polynomial by the second, then the second, $f_2(z)$, by the remainder from the first division, then this first remainder by the remainder from the second division, and so on, until we finally reach a division with a remainder equal to zero. *The last non-zero remainder is the highest common factor of the two given polynomials.* If this remainder does not contain z, the given polynomials are relatively prime. Thus, *finding the highest common factor amounts to division of the polynomials, arranged in decreasing powers of the variable.* Division of $f_1(z)$ and $f_2(z)$ by $D(z)$, their highest common factor, gives us relatively prime polynomials, one or both of which cannot contain z.

We see by comparing expansions (9) and (10) that the highest common factor of $f(z)$ and its derivative $f'(z)$ is:

$$D(z) = (z - z_1)^{k_1 - 1} (z - z_2)^{k_2 - 1} \dots (z - z_m)^{k_m - 1},$$

constant factors being omitted as of no account:

We obtain on dividing $f(z)$ by $D(z)$:

$$\frac{f(z)}{D(z)} = a_0 (z - z_1)(z - z_2) \dots (z - z_m),$$

i.e. *division of a polynomial $f(z)$ by the highest common factor of $f(z)$ and $f'(z)$ gives a polynomial, all the zeros of which are simple, and coincident with the zeros of $f(z)$.*

The operation of obtaining this polynomial is referred to as ridding $f(z)$ of multiple zeros. We see that this is achieved without needing to solve the equation $f(z) = 0$.

If $f(z)$ and $f'(z)$ are relatively prime, all the zeros of $f(z)$ are simple, and conversely.

189. Real polynomials. We now consider the polynomial with real coefficients:

$$f(z) = a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n ,$$

and we suppose that it has a complex zero $z = a + bi$ $(b \neq 0)$ of order of multiplicity k, i.e.

$$f(a + bi) = f'(a + bi) = \dots = f^{(k-1)}(a + bi) = 0 ;$$

$$f^{(k)}(a + bi) = A + Bi \neq 0.$$

We now replace all the quantities in $f(a + bi)$ and its derivatives by their conjugates. The real coefficients a_s will remain unchanged in this process, whilst $(a + bi)$ becomes $(a - bi)$, i.e. the polynomial $f(z)$ remains as before, except that its $z = a + bi$ is now replaced by $z = a - bi$. We know from [173] that the total effect of replacing complex numbers by their conjugates is to obtain the conjugate of their result, i.e. we get the conjugate of the polynomial. Thus:

$$f(a - bi) = f'(a - bi) = \dots = f^{(k-1)}(a - bi) = 0;$$

$$f^{(k)}(a - bi) = A - Bi \neq 0 .$$

i.e. *if a polynomial with real coefficients has a complex zero $z = a + bi$ of order of multiplicity k, it must have the conjugate zero $z = a - bi$ of the same multiplicity.*

The complex zeros of a polynomial with real coefficients are thus distributed in pairs of conjugate zeros. Suppose that the variable z only takes real values, and let these be denoted by x. In accordance with (3):

$$f(x) = a_0(x - z_1)(x - z_2) \dots (x - z_n) .$$

If there are imaginaries among the zeros z, the corresponding factors will also be imaginary. Multiplication of pairs of factors corresponding to pairs of conjugate zeros gives:

$$[x - (a + bi)]\,[x - (a - bi)] = [(x - a) - bi]\,[(x - a) + bi]$$
$$= (x - a)^2 + b^2 = x^2 + px + q\,,$$

where

$$p = -2a;\ q = a^2 + b^2\ (b \neq 0)\,.$$

Pairs of conjugate zeros give real factors of the second degree, so that we can state the proposition: *a polynomial with real coefficients can be factored into real factors of the first and second degrees.*

This expansion has the form:

$$f(x) = a_0(x - x_1)^{k_1}(x - x_2)^{k_2} \dots (x - x_r)^{k_r}(x^2 + p_1 x + q_1)^{l_1} \times$$
$$\times (x^2 + p_2 x + q_2)^{l_2} \dots (x^2 + p_t x + q_t)^{l_t}\,, \qquad (11)$$

where $x_1, x_2, \dots, x_r$ are real zeros of $f(x)$ of multiplicities $k_1, k_2 \dots, k_r$, and the factors of the second degree result from pairs of conjugate complex zeros of multiplicities $l_1, l_2, \dots, l_t$.

190. The relationship between the roots of an equation and its coefficients. Let the roots of the equation:

$$a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n = 0$$

be $z_1, z_2, \dots, z_n$ as before.

By (3), we have the identity:

$$a_0 z^n + a_1 z^{n-1} + \dots + a_{n-1} z + a_n = a_0(z - z_1)(z - z_2) \dots (z - z_n)\,.$$

We can apply to the right-hand side the familiar rule of elementary algebra for multiplying binomials that differ in their second terms, and write our identity as:

$$a_0 z^n + a_1 z^{n-1} + \dots + a_k z^{n-k} + \dots + a^n =$$
$$= a_0\,[z^n - S_1 z^{n-1} + S_2 z^{n-2} + \dots + (-1)^k S_k z^{n-k} + \dots + (-1)^n S_n]\,,$$

where S_k denotes the sum of all the possible products obtained by taking k of the z_s together ($s = 1, 2, \dots, n$). We find on comparing coefficients of the same powers of z:

$$S_1 = \frac{a_1}{a_0}\,;\quad S_2 = \frac{a_2}{a_1}\,;\ \dots\,;\ S_k = (-1)^k \frac{a_k}{a_0}\,;\ \dots\,;\ S_n = (-1)^n \frac{a_n}{a_0},$$

or in explicit form:

$$\left.\begin{aligned} z_1 + z_2 + \ldots + z_n &= -\frac{a_1}{a_0}\,; \\ z_1 z_2 + z_2 z_3 + \ldots + z_{n-1} z_n &= \frac{a_2}{a_0}\,; \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ z_1 z_2 \ldots z_n &= (-1)^n \frac{a_n}{a_0}\,. \end{aligned}\right\} \tag{12}$$

These formulae are a generalization of the familiar properties of the roots of quadratic equations to the case of equations of any degree. They enable us, incidentally, to construct an equation, given its roots.

191. Equations of the third degree. We shall not pay detailed attention to the actual calculation of the roots of algebraic equations. This subject is dealt with in text-books on approximate evaluation. We dwell only on the case of equations of the third degree, as well as sketching some methods of calculation that will be useful in future.

We start with an investigation of the third degree equation:

$$y^3 + a_1 y^2 + a_2 y + a_3 = 0. \tag{13}$$

We introduce a new variable x instead of y, putting

$$y = x + a.$$

Substitution of this in the left-hand side of (13) gives us:

$$x^3 + (3a + a_1)\,x^2 + (3a^2 + 2a_1 a + a_2)\,x + (a^3 + a_1 a^2 + a_2 a + a_3) = 0.$$

If we put $a = -(a_1/3)$, the term in x^2 falls out, so that the substitution

$$y = x - \frac{a_1}{3}$$

changes (13) to the form:

$$f(x) = x^3 + px + q = 0\,, \tag{14}$$

which does not contain x^2.

If p and q are real, (14) must have either three real roots, or one real and two conjugate complex roots [189]. We decide which of these cases is in question by taking the first derivative of the left-hand side of the equation:

$$f'(x) = 3x^2 + p.$$

If $p > 0$, $f'(x) > 0$, and $f(x)$ is always increasing; it will only have one real root, since its sign changes from $(-)$ to $(+)$ on passing from $x = -\infty$ to $x = +\infty$. Now take $p < 0$. It can easily be seen that $f(x)$ has a maximum at $x = -\sqrt{-p/3}$ and a minimum at $x = \sqrt{-(p/3)}$. Substitution of these

values of x in the expression for $f(x)$ gives us the respective maximum and minimum values of the function:

$$q \mp \frac{2p}{3}\sqrt{-\frac{p}{3}}.$$

If both these values have the same sign, i.e.

$$\left(q - \frac{2p}{3}\sqrt{-\frac{p}{3}}\right)\left(q + \frac{2p}{3}\sqrt{-\frac{p}{3}}\right) = q^2 + \frac{4p^3}{27} > 0$$

or

$$\frac{q^2}{4} + \frac{p^3}{27} > 0, \tag{15_1}$$

the equation has only one real root, lying in either the interval

$$\left(-\infty, -\sqrt{-\frac{p}{3}}\right) \quad \text{or in} \quad \left(+\sqrt{-\frac{p}{3}}, +\infty\right).$$

On the other hand, if the above maximum value of $f(x)$ is $(+)$ and the minimum is $(-)$, i.e.

$$\frac{q^2}{4} + \frac{p^3}{27} < 0, \tag{15_2}$$

the signs of $f(-\infty)$, $f(-\sqrt{-p/3})$, $f(+\sqrt{-p/3})$, $f(+\infty)$ are respectively $(-)$, $(+)$, $(-)$, $(+)$, and (14) has three real roots. We further remark that condition (15_1) is certainly satisfied when $p > 0$. We leave it to the reader to show that, for

$$\frac{q^2}{4} + \frac{p^3}{27} = 0, \tag{15_3}$$

(14) has a multiple root $\pm\sqrt{-p/3}$ and a root $3q/p$, taking $p \neq 0$, and so, from (15_3), $p < 0$. We have inequality (15_1) when $p = 0$ and $q \neq 0$, and equation (14) takes the form $x^3 + q = 0$, i.e. $x = \sqrt[3]{-q}$, whence it follows that (14) has one real root [175]. With $p = q = 0$, (14) becomes: $x^3 = 0$, and has the root of order three $x = 0$.

The results obtained are collected in the following table:

$x^3 + px + q = 0$	
$\frac{q^2}{4} + \frac{p^3}{27} > 0$	One real and two conjugate complex roots
$\frac{q^2}{4} + \frac{p^3}{27} < 0$	Three real and distinct roots
$\frac{q^2}{4} + \frac{p^3}{27} = 0$	Three real roots, one of which is repeated

Figure 182 shows the graphs of the function

$$y = x^3 + px + q$$

for the various assumptions regarding $(q^2/4 + p^3/27)$. The double root in case (15_3) corresponds to the point of contact of the curve with OX.

We now deduce a formula expressing the roots of equation (14) in terms of its coefficients. Whilst this formula is not convenient for practical calculations, a suitable practical method is derived from it in the next article, making use of trigonometric functions.

We introduce two new unknowns u and v in place of x, putting

$$x = u + v. \tag{16}$$

We substitute in (14):

$$(u + v)^3 + p(u + v) + q = 0\,,$$

$$u^3 + v^3 + (u + v)\,(3uv + p) + q = 0. \tag{17}$$

We subject u and v to the condition:

$$3uv + p = 0\,,$$

when (17) gives us:

$$u^3 + v^3 = -q.$$

The problem thus reduces to the solution of the two equations:

$$uv = -\frac{p}{3}\,;\; u^3 + v^3 = -q. \tag{18}$$

On raising both sides of the first equation to the third power, we have:

$$u^3v^3 = -\frac{p^3}{27}\,;\; u^3 + v^3 = -q,$$

so that u^3 and v^3 are roots of the quadratic equation:

$$z^2 + qz - \frac{p^3}{27} = 0,$$

i.e.

$$u = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}\,, \qquad v = \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}\,. \tag{19}$$

We finally obtain, from (16):

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}\,. \tag{20}$$

This formula for solving the cubic equation (14) is named after Cardan, a sixteenth century Italian mathematician.

We use R_1 and R_2 to denote, for brevity, the expressions under the cube root signs in (20):

$$x = \sqrt[3]{R_1} + \sqrt[3]{R_2},$$

Each of the cube roots has three distinct values [175], so that the formula written gives in general nine distinct values of x, only three of which can be roots of (14). The extra values of x are due to the fact that we cubed the first of equations (18). We can only consider those values, for which u and v are connected by the first of equations (18), i.e. we *must only take the values of the cube roots whose products are equal to* $(-p/3)$.

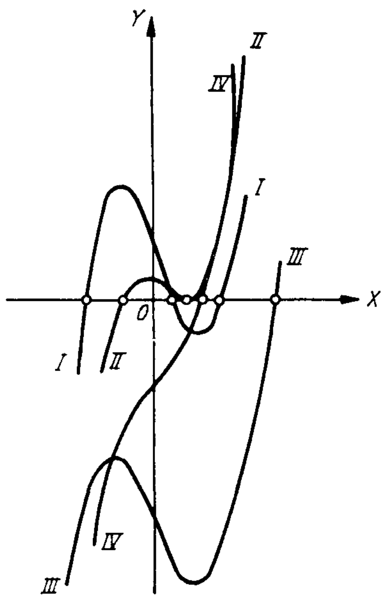

FIG. 182

We denote by ε one of the values of the cube root of unity:

$$\varepsilon = \cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3} = -\frac{1}{2} + \frac{\sqrt{3}}{2}i,$$

$$\varepsilon^2 = \cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3} = -\frac{1}{2} - \frac{\sqrt{3}}{2}i$$

and we let $\sqrt[3]{R_1}$ and $\sqrt[3]{R_2}$ denote values of the roots that satisfy the above condition. On multiplying these by ε and ε^2 we obtain all three values of the root [175].

If we recall that $\varepsilon^3 = 1$, we get the following expressions for the roots of equation (14), in which p and q are taken to be any complex numbers:

$$x_1 = \sqrt[3]{R_1} + \sqrt[3]{R_2}\,;\ x_2 = \varepsilon\sqrt[3]{R_1} + \varepsilon^2\sqrt[3]{R_2}\,;\ x_3 = \varepsilon^2\sqrt[3]{R_1} + \varepsilon\sqrt[3]{R_2}\,. \tag{21}$$

192. The trigonometric form of solution of cubic equations. We assume that the coefficients p and q in equation (14) are real. It has already been mentioned that Cardan's formula is inconvenient for calculating the roots, and we now deduce a more practical formula. We consider four separate cases.

1. $$\frac{q^2}{4} + \frac{p^3}{27} < 0.$$

It follows from this that $p < 0$. The expressions R_1 and R_2 under the radicals in (20) will be imaginary, yet in spite of this, all three roots of the equation will be real, as we know [191].

We put:

$$-\frac{q}{2} \pm \sqrt{\frac{q^2}{4} + \frac{p^3}{27}} = -\frac{q}{2} \pm i\sqrt{-\frac{q^2}{4} - \frac{p^3}{27}} = r(\cos\varphi \pm i\sin\varphi),$$

whence [171]:

$$r = \sqrt{-\frac{p^3}{27}}; \quad \cos\varphi = -\frac{q}{2r}. \tag{22}$$

We have by Cardan's formula:

$$x = \sqrt[3]{r}\left(\cos\frac{\varphi + 2k\pi}{3} + i\sin\frac{\varphi + 2k\pi}{3}\right) + $$
$$+ \sqrt[3]{r}\left(\cos\frac{\varphi + 2k\pi}{3} - i\sin\frac{\varphi + 2k\pi}{3}\right) \qquad (k = 0,\ 1\ 2).$$

If we take the same values of k in both terms, we obtain for the products of these terms the positive number: $\sqrt[3]{r^2} = -p/3$.

We have finally:

$$x = 2\sqrt[3]{r}\cos\frac{\varphi + 2k\pi}{3} \qquad (k = 0,\ 1,\ 2), \tag{23}$$

where r and φ are defined by (22), and where it can easily be shown that, if we take different φ satisfying the second of equations (22), we get the same set of roots from (23).

2. $$\frac{q^2}{4} + \frac{p^3}{27} > 0 \quad \text{and} \quad p < 0.$$

Equation (14) has one real, and two conjugate complex, roots [191], whilst it follows from the conditions written that $-(p^3/27) < (q^2/4)$. We introduce the auxiliary angle ω, putting:

$$\sqrt{-\frac{p^3}{27}} = \frac{q}{2}\sin\omega. \tag{24_1}$$

This gives us:

$$\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{-\frac{q}{2} + \frac{q}{2}\cos\omega} = -\sqrt{-\frac{p}{3}}\sqrt[3]{\tan\frac{1}{2}\omega},$$

$$\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{-\frac{q}{2} - \frac{q}{2}\cos\omega} = -\sqrt{-\frac{p}{3}}\sqrt[3]{\cot\frac{1}{2}\omega},$$

since, by (24_1):

$$\sqrt{-\frac{p}{3}} = \sqrt[3]{\frac{1}{2}q\sin\omega}.$$

We finally introduce the angle φ, where

$$\tan\varphi = \sqrt[3]{\tan\frac{1}{2}\omega}, \tag{24_2}$$

and get the following expression for the real root:

$$x_1 = -\sqrt{-\frac{p}{3}}(\tan\varphi + \cot\varphi) = -\frac{2\sqrt{-\frac{p}{3}}}{\sin 2\varphi}. \tag{25_1}$$

We suggest that the reader make use of (21) to show that the imaginary roots will have the form:

$$\frac{1}{\sin 2\varphi}\sqrt{-\frac{p}{3}} \pm i\sqrt{-p}\cot 2\varphi. \tag{25_2}$$

3. $$\frac{q^2}{4} + \frac{p^3}{27} > 0 \quad \text{and} \quad p > 0.$$

In this, as in the previous case, (14) has one real and two conjugate complex roots. The value here of $\sqrt{p^3/27}$ can be both greater and less than $|q/2|$, and we take an angle ω defined by the following, instead of by (24_1):

$$\sqrt{\frac{p^3}{27}} = \frac{1}{2}q\tan\omega. \tag{26_1}$$

This gives:

$$\sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{\frac{q\sin^2\frac{1}{2}\omega}{\cos\omega}} = \sqrt{\frac{p}{3}}\sqrt[3]{\tan\frac{1}{2}\omega},$$

$$\sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} = \sqrt[3]{\frac{q\cos^2\frac{1}{2}\omega}{\cos\omega}} = \sqrt{\frac{p}{3}}\sqrt[3]{\cot\frac{1}{2}\omega}.$$

We introduce a new angle φ defined by:

$$\tan\varphi = \sqrt[3]{\tan\frac{1}{2}\omega}, \tag{26_2}$$

and finally have:

$$x_1 = \sqrt{\frac{p}{3}}(\tan\varphi - \cot\varphi) = -2\sqrt{\frac{p}{3}}\cot 2\varphi. \tag{27_1}$$

The imaginary roots will be:

$$\sqrt{\frac{p}{3}}\cot 2\varphi \pm \frac{i\sqrt{p}}{\sin 2\varphi}. \tag{27_2}$$

4. $$\frac{q^2}{4} + \frac{p^3}{27} = 0.$$

Equation (14) has a repeated root, and here, as in the case of $p = 0$, the solution presents no difficulty.

The roots of a cubic equation can be calculated with considerable accuracy by using the trigonometric formulae deduced, together with logarithmic tables.

Examples. 1.

$$x^3 + 9x^2 + 23x + 14 = 0.$$

We put $x = y - 3$ and reduce the equation to the form:

$$y^3 - 4y - 1 = 0,$$

which has three real roots [191].

We obtain $\cos\varphi$ from (22), and having found φ itself, we determine the roots from (23):

$\cos\varphi = \frac{\sqrt{27}}{16}$; $\log_{10}\cos\varphi = \bar{1}.51156$
$\varphi = 71°2'56''$
$\frac{\varphi_1}{3} = 23°40'59''$; $\frac{\varphi_2}{3} = 143°40'59''$; $\frac{\varphi_3}{3} = 263°40'59''$;
$\log_{10}\frac{4}{\sqrt{3}} = 0.36350$
$\log_{10} y_1 = 0.32529$; $\log_{10}(-y_2) = 0.26970$; $\log_{10}(-y_3) = \bar{1}.40501$
$y_1 = 2.1149$; $y_2 = -1.8608$; $y_3 = -0.2541$
$x_1 = -0.8851$; $x_2 = -4.8608$; $x_3 = -3.2541$

2.

$$x^3 - 3x + 5 = 0.$$

We determine ω from (24_1) and φ from (24_2), then calculate the roots from (25_1) and (25_2):

$\log_{10}\sin\omega = \bar{1}.60206$; $\omega = 23°34'11''$; $\frac{1}{2}\omega = 11°47'20''$
$\log_{10}\tan\varphi = \bar{1}.77009$; $\varphi = 30°29'47''$; $2\varphi = 60°59'34''$
$\log_{10}\frac{1}{\sin 2\varphi} = 0.05821$; $\frac{1}{\sin 2\varphi} = 1.1434$
$\log_{10}\sqrt{-p}\cot 2\varphi = \bar{1}.98244$; $\sqrt{-p}\cot 2\varphi = 0.96037$
$x_1 = -2.2868$; $x_2, x_3 = 1.1434 \pm 0.96037\,i$

193. The method of successive approximations. It is possible in many cases conveniently to improve upon an approximate value x_0 of a required root ξ, when x_0 is known to a certain number of decimal places. One such method of *correcting* an approximate value of a root is that of *successive approximations*. As is explained later, this method is suitable for transcendental, as well as algebraic, equations.

We suppose that the equation

$$f(x) = 0 \tag{28}$$

has been rewritten in the form:

$$f_1(x) = f_2(x) , \tag{29}$$

where $f_1(x)$ is such that the equation

$$f_1(x) = m$$

has one real root, easily calculable with great accuracy, for any real m. Evaluation of the roots of equation (29) by the method of successive ap-

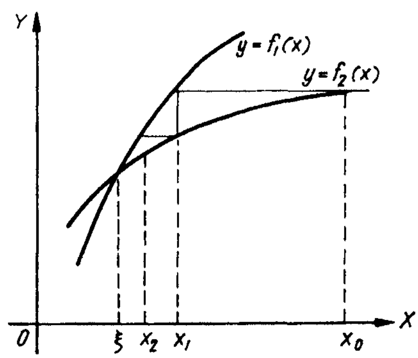

FIG. 183

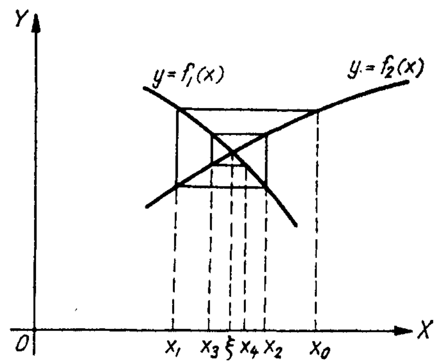

FIG. 184

proximations proceeds as follows: we substitute an approximate value x_0 of the required root in the right-hand side of (29) and find a second approximation x_1 to the root from the equation:

$$f_1(x) = f_2(x_0) .$$

We substitute x_1 in the right-hand side of (29) and get the next approximation x_2 by solving the equation $f_1(x) = f_2(x_1)$, and so on. We thus find a sequence of values:

$$x_0, x_1, x_2, \ldots, x_n, \ldots , \tag{30}$$

with

$$f_1(x_1) = f_2(x_0);\ f_1(x_2) = f_2(x_1);\ \ldots,\ f_1(x_n) = f_2(x_{n-1});\ \ldots \tag{31}$$

The geometrical significance of the approximations obtained is easily shown. The required root is the abscissa of the point of intersection of the curves:

$$y = f_1(x) \tag{32$_1$}$$

and

$$y = f_2(x). \tag{32$_2$}$$

The curves are illustrated in Figs. 183 and 184, the derivatives $f_1'(x)$ and $f_2'(x)$ having the same sign at the point of intersection in the case of Fig. 183, and different signs in Fig. 184, whilst in both cases

$$|f_2'(\xi)| < |f_1'(\xi)|.$$

The following construction corresponds to equations (31): we draw the straight line $x = x_0$, parallel to OY, to its point of intersection (x_0, y_0) with the curve (32_2); we draw through this point the line $y = y_0$, parallel to OX, to its point of intersection (x_1, y_0) with curve (32_1); now we draw through (x_1, y_0) the line $x = x_1$, parallel to OY, to its point of intersection (x_1, y_1) with the curve (32_2); next we draw through this last point the line $y = y_1$ to its intersection at (x_2, y_1) with the curve (32_1), and so on. The abscissae of the points of intersection give us the sequence (30).

If the first approximation is taken sufficiently close to ξ, this sequence tends to ξ as a limit, as is clear from the figure; in the case where $f_1'(\xi)$ and $f_2'(\xi)$ have the same sign, a step-line tending to ξ is obtained (Fig. 183) whereas when $f_1'(\xi)$ and $f_2'(\xi)$ have different signs, we get a form of rectangular spiral tending to ξ (Fig. 184). We shall not go into the condition for, and rigorous proof of, the fact that sequence (30) tends to ξ as a limit. This can be observed directly from the figure in many cases.

The method described is particularly convenient to apply in the case when equation (29) has the form:

$$x = f_2(x).$$

Let ξ be the root of this equation, and let an approximate value of it,

$$x_0 = \xi + h,$$

be known.

The sequence of approximations will be:

$$x_1 = f_2(x_0);\ x_2 = f_2(x_1),\ \ldots;\ x_n = f_2(x_{n-1})\ \ldots$$

It can be shown that in fact $x_n \to \xi$ as $n \to \infty$, if the function $f_2(x)$ has a derivative $f_2'(x)$ which satisfies the condition:

$$|f_2'(x)| \leqslant q < 1,$$

when

$$\xi - h \leqslant x \leqslant \xi + h.$$

Examples. **1.** We take the equation

$$x^5 - x - 0.2 = 0. \tag{33}$$

Its real roots are the points of intersection of (Fig. 185):

$$y = x^5, \tag{34$_1$}$$

$$y = x + 0.2 \tag{34$_2$}$$

and (33) has one positive and two negative roots, as can be seen from Fig. 185.

At the points of intersection A and B, corresponding respectively to the positive root and to the negative root with the greater absolute value, the absolute value of the slope of the straight line (34_2) is less than the absolute value of the slope of the tangent to curve (34_1) i.e. calculation of these roots by the method of successive approximations must proceed with equation (33) set in the form:

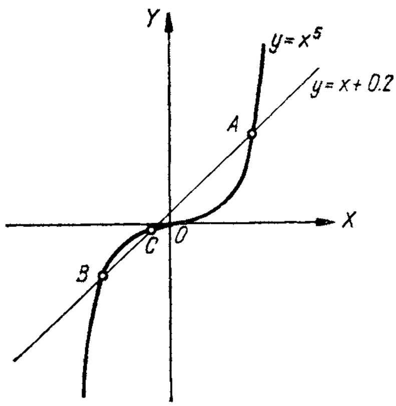

FIG. 185

$$x^5 = x + 0.2.$$

We take $x_0 = 1$ as a first approximation for calculating the positive root, and obtain the table:

$\sqrt[5]{x_n + 0.2}$	$x_n + 0.2$
	1.2
$x_1 = 1.037$	1.237
$x_2 = 1.0434$	1.2434
$x_3 = 1.0445$	1.2445
$x_4 = 1.04472$	

With x_4, the required root is obtained to an accuracy of four decimal places.

We evaluate the negative root with the greater absolute value by taking $x_0 = -1$ as a first approximation.

$\sqrt[5]{x_n + 0.2}$	$x_n + 0.2$
	-0.8
$x_1 = -0.956$	-0.756
$x_2 = -0.9456$	-0.7456
$x_3 = -0.9430$	-0.7430
$x_4 = -0.9423$	-0.7423
$x_5 = -0.94214$	-0.74214
$x_6 = -0.94210$	

The error in this case does not exceed 2×10^{-5}.

At C, corresponding to the negative root with the smaller absolute value, the absolute value of the slope of the tangent to the curve (34_1) is less than unity, so that use of the method of successive approximations implies setting (33) in the form:

$$x = x^5 - 0.2.$$

We take $x_0 = 0$ as a first approximation, and obtain:

$x_n^5 - 0.2$	x_n^5
	-0
$x_1 = -0.2$	-0.00032
$x_2 = -0.20032$	

The approximation x_2 gives the root to an accuracy of five places. The approximation to the root proceeds by a step-line in all three cases, as illustrated in Fig. 183, and as is easily verified from Fig. 185; also, x_n tends monotonically to the required root with increase of n in all three cases.

2.

$$x = \tan x. \tag{35}$$

The roots of this equation are the points of intersection (Fig. 186) of

$$y = x, \; y = \tan x,$$

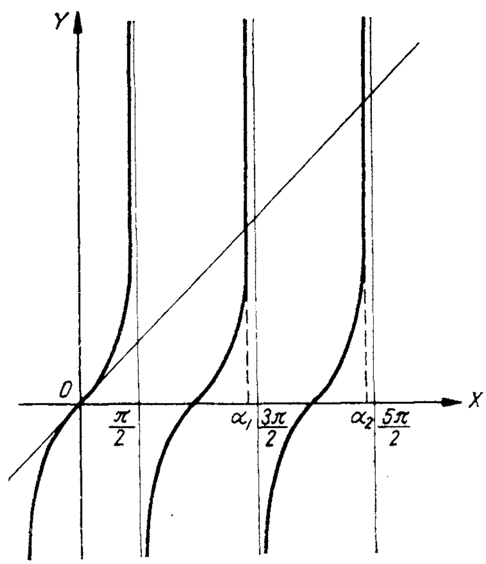

Fig. 186

and, as can be seen from the figure, the equation has one root in each of the intervals

$$\left[(2n-1)\frac{\pi}{2}, \quad (2n+1)\frac{\pi}{2}\right]$$

$$(n = 0, \pm 1, \pm 2, \ldots).$$

The following approximation will hold for positive roots:

$$a_n \sim (2n+1)\frac{\pi}{2},$$

where a_n denotes the nth positive root of equation (35).

We evaluate the root a_1, close to $3\pi/2$. We apply the method of successive approximations after rewriting (35) in the form:

$$x = \text{arc tan}\, x,$$

and we take $x_0 = 3\pi/2$ as the first approximation.

When calculating the sequence of approximations

$$x_n = \text{arc tan}\, x_{n-1},$$

we must always take the value of the arc tangent lying in the third quadrant. We obtain, using logarithmic tables and expressing the arcs in radian measure:

$$x_0 = 4.7124; \qquad x_1 = 4.5033;$$

$$x_2 = 4.4938; \qquad x_3 = 4.4935.$$

194. Newton's method. The process of successive approximations shown in Figs. 183 and 184 consists in approximating to the required root by means of lines parallel to the coordinate axes. We now give some other similar processes, in which use is made of straight lines inclined to the axes. Newton's method is one of these.

Let x_0' and x_0 be approximate values of the root ξ of the equation

$$f(x) = 0 \tag{36}$$

and let ξ be the only root of the equation in the interval (x_0', x_0). Graphs of

$$y = f(x)$$

are illustrated in Figs. 187 and 188.

The abscissae of points N and P are the approximations x_0' and x_0 to the root ξ, which is given by the abscissa of A. The tangent PQ_1 to the curve is drawn at the point P, and from the point of intersection Q_1 of this tangent with the axis of abscissae we draw the ordinate $\overline{Q_1Q}$ of the curve; at the point Q we draw the tangent QR_1 to the curve and from the point R_1 draw the ordinate $\overline{R_1R}$ of the curve and so on.

It is clear from the figure that the points $P_1, Q_1, R_1 \ldots$ tend to A, so that their abscissae, $x_0\, x_1, x_2 \ldots$ form a sequence of approximations to ξ. We deduce a formula, expressing x_n in terms of x_{n-1}.

The equation of the tangent PQ_1 is:

$$Y - f(x_0) = f'(x_0)\,(X - x_0).$$

We find the abscissa of Q_1 by substituting $Y = 0$:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}, \tag{37}$$

and in general:

$$x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})} \tag{38}$$

$$(n = 1, 2, 3 \ldots).$$

We have simply observed by inspection of the figures that the x_n are approximations to the root ξ, the figures being drawn for the case when $f(x)$ is monotonic and remains convex (or concave) in the interval (x_0', x_0),

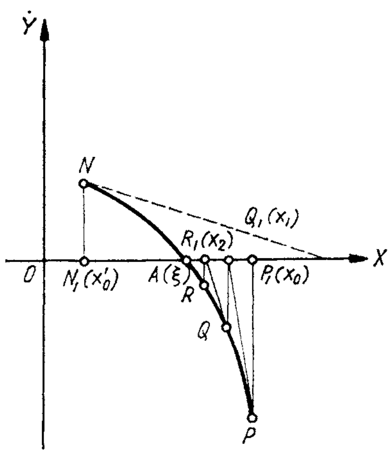

FIG. 187

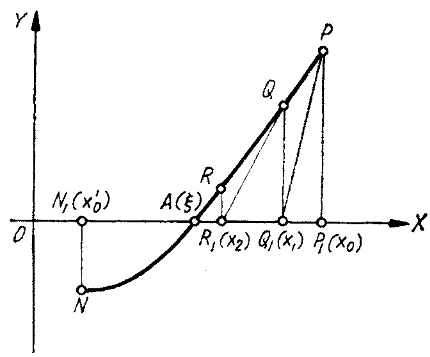

FIG. 188

in other words, when $f'(x)$ and $f''(x)$ preserve their sign in the interval [57, 71]. We shall not dwell on a rigorous analytic proof.

We remark that if we applied Newton's method to the end x_0' instead of to x_0, approximations to the root would not be obtained, as is indicated by the tangent drawn with a broken line. In the case of Fig. 188, the curve becomes concave in the region of positive ordinates, i.e. $f''(x) > 0$, and Newton's method has to be applied at the end where $f(x) > 0$, as we have seen. It follows from Fig. 187 that, with $f''(x) < 0$, Newton's method must be applied to the end where the ordinate $f(x) < 0$. We thus arrive at the following rule: *if $f'(x)$ and $f''(x)$ do not vanish in the interval (x_0', x_0), whilst the ordinates $f(x_0')$ and $f(x_0)$ have different signs, application of Newton's method to the end of the interval at which $f(x)$ and $f''(x)$ have the same sign gives us a sequence of approximations to the unique root of equation* (36) *contained in the interval.*

195. The method of simple interpolation. We mention one further method of approximate evaluation of roots. We draw a straight line through the ends N, P of the arc of the curve. The abscissa of the intersection B of this

line with axis OX gives an approximate value of the root (Fig. 189). As before, let x_0', x_0 be the abscissae of the ends of the interval. The equation of NP is:

$$\frac{Y - f(x_0)}{f(x_0') - f(x_0)} = \frac{X - x_0}{x_0' - x_0}.$$

We find the expression for the abscissa of B by putting $Y = 0$:

$$\frac{x_0' f(x_0) - x_0 f(x_0')}{f(x_0) - f(x_0')}$$

or

$$x_0' - \frac{(x_0 - x_0') f(x_0')}{f(x_0) - f(x_0')}, \tag{39}$$

or

$$x_0 - \frac{(x_0 - x_0') f(x_0)}{f(x_0) - f(x_0')}.$$

Replacing the segment of curve by the straight line passing through its ends is equivalent to replacing the function $f(x)$ in the interval by a first degree polynomial having the same end values as $f(x)$; or alternatively, and this amounts to the same thing, it is equivalent to assuming that the, variation of $f(x)$ in the interval is proportional to the variation of x. This method is usually referred to as *simple interpolation*, and is applied, for instance, when using logarithmic tables (proportional parts). It is also sometimes called the *rule of false position (regula falsi)*.

If simple interpolation is used simultaneously with Newton's method the possibility arises of approximating both limits x_0' and x_0 to the root ξ.

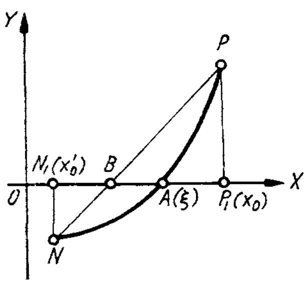

FIG. 189

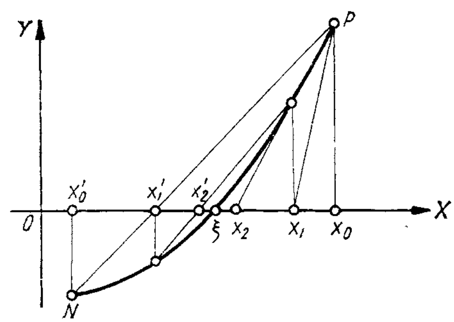

FIG. 190

Suppose, for example, that $f(x)$ and $f''(x)$ have the same sign at the end x_0, so that Newton's method has to be used with reference to this end.

The use of both methods gives us two new approximate values (Fig. 190):

$$x_0' = \frac{x_0' f(x_0) - x_0 f(x_0')}{f(x_0) - f(x_0')};$$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}.$$

We can again apply the same formulae to the approximate values x_1' and x_1, and obtain the new values:

$$x_2' = \frac{x_1' f(x_1) - x_1 f(x_1')}{f(x_1) - f(x_1')} ;$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} .$$

We thus obtain two series of values:

$$x_0', x_1', x_2', \ldots, x_n', \ldots ,$$

and

$$x_0, x_1, x_2, \ldots, x_n, \ldots ,$$

approximating to the root ξ from the left and from the right.

If x_n' and x_n coincide to several decimal places, the root ξ, lying between x_n' and x_n, must be obtained to the same number of decimal places.

Example. The equation

$$f(x) = x^5 - x - 0.2 = 0 ,$$

which we considered in Example 1 of [193], has one positive root in the inter val $1 < x < 1.1$, and

$$f'(x) = 5x^4 - 1 \quad \text{and} \quad f''(x) = 20x^3$$

do not change sign in this interval. We can therefore put

$$x_0' = 1; \quad x_0 = 1.1 .$$

We calculate the values of $f(x)$:

$$f(1) = -0.2; \; f(1.1) = 0.31051 ,$$

from which it is clear that $f(x)$ and $f''(x)$ have the same (plus) sign at the right-hand end ($x_0 = 1.1$), so that it is in regard to this end that Newton's method must be applied.

We firstly evaluate $f'(x)$ at the right-hand end:

$$f'(1.1) = 6.3205 .$$

We have by (37) and (39):

$$x_1' = 1 + \frac{0.1 \times 0.2}{0.51051} = 1.039 ,$$

$$x_1 = 1.1 - \frac{0.31051}{6.3205} = 1.051 .$$

We evaluate for the next approximation:

$$f(1.039) = -0.0282; \; f(1.051) = 0.0313; \; f'(1.051) = 5.1005 ,$$

whence:

$$x_2' = 1.039 + \frac{0.012 \times 0.0282}{0.0595} = 1.04469\,,$$

$$x_2 = 1.051 - \frac{0.0313}{5.1005} = 1.04487\,,$$

which gives the value of the root to an accuracy of five places within two units [193]:

$$1.04469 < \xi < 1.04487.$$

§ 19. Integration of various functions

196. Reduction of rational fractions to partial fractions. We described above a number of methods of evaluating indefinite integrals. We supplement this description in the present section and give it a more systematic character. The first problem is that of integrating *rational fractions*, i.e. the quotients of two polynomials. Before we turn to the solution of this problem, we establish a formula which represents a rational fraction as the sum of several fractions of the simplest type. This is known as *the reduction of a rational fraction to partial fractions.*

Let us be given the rational fraction:

$$\frac{F(x)}{f(x)}.$$

If it is an improper fraction, i.e. the degree of the numerator is not lower than that of the denominator, we can divide out an integral part consisting of the polynomial $Q(x)$ and write the fraction in the form:

$$\frac{F(x)}{f(x)} = Q(x) + \frac{\varphi(x)}{f(x)}\,, \tag{1}$$

where $\varphi(x)/f(x)$ is now a proper fraction, the numerator of which is of lower degree than the denominator. We shall further assume that this latter fraction is irreducible, i.e. that the numerator and denominator are relatively prime [188].

Let $x = a$ be a zero of order k of the denominator:

$$f(x) = (x-a)^k f_1(x) \text{ and } f_1(a) \neq 0.$$

We show that the fraction can be written in the form of a sum:

$$\frac{\varphi(x)}{(x-a)^k f_1(x)} = \frac{A}{(x-a)^k} + \frac{\varphi_1(x)}{(x-a)^{k-1} f_1(x)}\,, \tag{2}$$

where A is a constant and the second term on the right-hand side is a proper fraction.

We form the difference:

$$\frac{\varphi(x)}{(x-a)^k f_1(x)} - \frac{A}{(x-a)^k} = \frac{\varphi(x) - Af_1(x)}{(x-a)^k f_1(x)}$$

and define A in such a way that the numerator of the fraction on the right-hand side of the equation written is divisible by $(x - a)$ [184]:

$$\varphi(a) - Af_1(a) = 0\,,$$

whence

$$A = \frac{\varphi(a)}{f_1(a)} \qquad (f_1(a) \neq 0).$$

We can cancel $(x - a)$ in the right-hand side just mentioned with this choice of A, and we thus arrive at identity (2). This shows that separation of the term of the form $A/(x-a)^k$, which is known as a partial fraction, enables us to lower the power of the factor $(x - a)^k$ appearing in the denominator by at least unity.

Suppose that the denominator can be written in the factor form:

$$f(x) = (x-a_1)^{k_1}(x-a_2)^{k_2} \ldots (x-a_m)^{k_m}.$$

We do not write a constant factor, since this can be divided into the numerator. Repeated application of the above rule for separating out the partial fraction gives us the reduction of a proper rational fraction to partial fractions:

$$\begin{aligned}\frac{\varphi(x)}{f(x)} = {} & \frac{A_{k_1}^{(1)}}{(x-a_1)^{k_1}} + \frac{A_{k_1-1}^{(1)}}{(x-a_1)^{k_1-1}} + \ldots + \frac{A_1^{(1)}}{x-a_1} + \\ & + \frac{A_{k_2}^{(2)}}{(x-a_2)^{k_2}} + \frac{A_{k_2-1}^{(2)}}{(x-a_2)^{k_2-1}} + \ldots + \frac{A_1^{(2)}}{x-a_2} + \\ & + \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots + \\ & + \frac{A_{k_m}^{(m)}}{(x-a_m)^{k_m}} + \frac{A_{k_m-1}^{(m)}}{(x-a_m)^{k_m-1}} + \ldots + \frac{A_1^{(m)}}{x-a_m}. \end{aligned} \tag{3}$$

We now give some methods of determining the coefficients appearing on the right-hand side of the identity written. If we clear this of the denominators, we arrive at an identity between two polynomials, and on equating corresponding coefficients, we get a system of linear

equations defining the required coefficients. As mentioned above [185], we call this the method of undetermined *coefficients*.

A different approach can be used, in which we assign particular values to the variable x in the identity between the two polynomials mentioned above. This *substitution method* can also be used as a preliminary to repeated differentiation of the identity (3).

We shall not dwell on the proof of the fact that expansion (3) is unique, i.e. its coefficients have fully defined values, independent of the method of expansion. We give examples later of using these methods for finding the unknown coefficients of the expansion.

Even when the polynomials $\varphi(x)$ and $f(x)$ are real, the right-hand side of (3) may still contain imaginary terms, resulting from imaginary zeros of the denominator. We shall mention another expansion of a rational fraction which is free from this defect, though here we confine ourselves to the case when the denominator of the fraction has only simple zeros, since this case is of the most value in applications.

The sum of partial fractions corresponding to a pair of conjugate complex zeros of the denominator, $x = a \pm bi$, is

$$\frac{A + Bi}{x - a - bi} + \frac{A - Bi}{x - a + bi}\,.$$

If we reduce these fractions to a common denominator, we get a partial fraction of the form:

$$\frac{Mx + N}{x^2 + px + q} \qquad (p = -2a\,;\; q = a^2 + b^2).$$

We can thus reduce the real rational fraction in this case to real partial fractions:

$$\begin{aligned}\frac{\varphi(x)}{f(x)} &= \frac{A_1}{x - a_1} + \frac{A_2}{x - a_2} + \dots + \frac{A_r}{x - a_r} \\ &+ \frac{M_1x + N_1}{x^2 + p_1x + q_1} + \frac{M_2x + N_2}{x_2 + p_2x + q_2} + \dots + \frac{M_sx + N_s}{x^2 + p_sx + q_s}\,, \qquad (4)\end{aligned}$$

where the fractions in the first line correspond to real zeros of the denominator, and those in the second line to pairs of conjugate complex roots.

197. Integration of rational fractions. Integration of a rational fraction leads by (1) to integration of a polynomial, which gives another polynomial, and to integration of a proper rational fraction, which we now consider.

If the denominator only has simple zeros, the problem reduces, by (4), to the two types of integral:

1. $$\int \frac{A}{x-a}\,dx = A\log(x-a) + C$$

and

2. $$\int \frac{Mx+N}{x^2+px+q}\,dx.$$

Recalling what was said in [92], we get a result of the form:

$$\int \frac{Mx+N}{x^2+px+q}\,dx = \lambda \log(x^2+px+q) + \mu \operatorname{arc\,tan} \frac{2x+p}{\sqrt{4q-p^2}} + C.$$

The integral is expressed here in terms of a logarithm and an arc tangent.

We now take the case when the denominator of the proper rational fraction contains multiple zeros. We return to expansion (3). Any imaginary numbers that may appear in it will only play an intermediary role in the subsequent working and will vanish in the final result.

Integration of a partial fraction in which the denominator has a degree higher than the first yields another rational fraction:

$$\int \frac{A^{(i)}_{k_i-s}}{(x-a_i)^{k_i-s}}\,dx = \frac{A^{(i)}_{k_i-s}}{(1-k_i+s)(x-a_i)^{k_i-s-1}} + C \quad (k_i - s > 1).$$

The sum of the fractions obtained after integration gives the *algebraic part* of the integral, which, on reduction to a common denominator, is clearly a proper fraction of the type:

$$\frac{\omega(x)}{(x-h)^{k_1-1}(x-a_2)^{k_2-1}\dots(x-a_m)^{k_m-1}}.$$

The numerator $\omega(x)$ is a polynomial of degree lower by at least unity than the degree of the denominator, whilst the denominator consists of the highest common factor $D(x)$ of the denominator $f(x)$ of the integrand and its first derivative $f'(x)$ [188].

The sum of the fractions not yet integrated:

$$\frac{A^{(1)}_1}{x-a_1} + \frac{A^{(2)}_1}{x-a_2} + \dots + \frac{A^{(m)}_1}{x-a_m}$$

gives on reduction to a common denominator a proper fraction of the form:

$$\frac{\omega_1(x)}{(x-a_1)(x-a_2)\dots(x-a_m)},$$

where $\omega_1(x)$ is a polynomial of degree lower by at least unity than the degree of the denominator, whilst the denominator consists of the quotient $D_1(x)$ of $f(x)$ divided by $D(x)$. We thus obtain the Ostrogradskii–Hermite formula:

$$\int \frac{\varphi(x)}{f(x)}\,\mathrm{d}x = \frac{\omega(x)}{D(x)} + \int \frac{\omega_1(x)}{D_1(x)}\,\mathrm{d}x\,. \tag{5}$$

We can find $D(x)$ and $D_1(x)$ without knowing the zeros of $f(x)$ [188]. We now show how to determine the coefficients of the polynomials $\omega(x)$ and $\omega_1(x)$, the degrees of which we can take as one less than the degrees of the corresponding denominators. We get rid of the integral signs by differentiating equation (5). We get rid of the denominators in the identity thus obtained, and get an identity between two polynomials; on applying to this latter the method of undetermined coefficients or the substitution method, we can find the coefficients of $\omega(x)$ and $\omega_1(x)$.

The Ostrogradskii–Hermite formula thus gives the algebraic part of the integral of a proper rational fraction, even in the case when the zeros of the denominator are unknown. The denominator of the fraction under the integral sign on the right-hand side of equation (5) contains only simple zeros, and we can evaluate the integral by reducing to partial fractions, the result being expressed in terms of logarithms and arc tangents, as we have seen. We have to know the zeros of $D_1(x)$ for this last operation.

Example. Using the Ostrogradskii-Hermite formula:

$$\int \frac{\mathrm{d}x}{(x^3+1)^2} = \frac{\alpha x^2 + \beta x + \gamma}{x^3+1} + \int \frac{\delta x^2 + \varepsilon x + \eta}{x^3+1}\,\mathrm{d}x\,.$$

We differentiate with respect to x:

$$\frac{1}{(x^3+1)^2} = \frac{(2\alpha x + \beta)(x^3+1) - 3x^2(\alpha x^2 + \beta x + \gamma)}{(x^3+1)^2} + \frac{\delta x^2 + \varepsilon x + \eta}{x^3+1}$$

and obtain, on clearing fractions:

$$1 = (2\alpha x + \beta)(x^3+1) - 3x^2(\alpha x^2 + \beta x + \gamma) + (\delta x^2 + \varepsilon x + \eta)(x^3+1).$$

We compare coefficients of x^5 and find $\delta = 0$, then compare coefficients of x^2 and find $\gamma = 0$. We substitute $\gamma = \delta = 0$ in the identity written and compare coefficients of the remaining powers, and find:

$$\varepsilon - \alpha = 0;\ \eta - 2\beta = 0;\ 2\alpha + \varepsilon = 0;\ \beta + \eta = 1\,,$$

whence finally:

$$\alpha = \gamma = \delta = \varepsilon = 0\,; \quad \beta = \frac{1}{3}\,; \quad \eta = \frac{2}{3},$$

so that

$$\int \frac{\mathrm{d}x}{(x^3+1)^2} = \frac{x}{3\,(x^3+1)} + \frac{2}{3}\int \frac{\mathrm{d}x}{x^3+1}.$$

The last integral is evaluated by reducing to partial fractions:

$$\frac{1}{x^3+1} = \frac{A}{x+1} + \frac{Mx+N}{x^2-x+1}.$$

We clear fractions:

$$1 = A(x^2 - x + 1) + (Mx + N)\,(x + 1).$$

Substituting $x = -1$ gives $A = 1/3$, then comparing coefficients of x^2 and the absolute terms gives:

$$M = -\frac{1}{3}\,; \quad N = \frac{2}{3},$$

and consequently,

$$\frac{1}{x^3+1} = \frac{1}{3\,(x+1)} - \frac{x-2}{3\,(x^2-x+1)}.$$

We finally get:

$$\int \frac{\mathrm{d}x}{x^3+1} = \frac{1}{3}\int \frac{\mathrm{d}x}{x+1} - \frac{1}{3}\int \frac{x-2}{x^2-x+1}\,\mathrm{d}x =$$

$$= \frac{1}{3}\log(x+1) - \frac{1}{6}\log(x^2-x+1) + \frac{1}{\sqrt{3}}\,\mathrm{arc\,tan}\,\frac{2x-1}{\sqrt{3}} + C,$$

whence

$$\int \frac{\mathrm{d}x}{(x^3+1)^2} = \frac{x}{3\,(x^3+1)} + \frac{2}{9}\log(x+1) - \frac{1}{9}\log(x^2-x+1)$$
$$+ \frac{2}{3\sqrt{3}}\,\mathrm{arc\,tan}\,\frac{2x-1}{\sqrt{3}} + C.$$

198. Integration of expressions containing radicals. We consider some other types of integral, which reduce to integrals of rational fractions.

1. The integral

$$\int R\left[x, \left(\frac{ax+b}{cx+d}\right)^{\lambda}, \left(\frac{ax+b}{cx+d}\right)^{\mu}, \ldots\right]\mathrm{d}x, \tag{6}$$

where R is a rational function of its arguments, i.e. the quotient of polynomials in these arguments, whilst λ, μ, ... are rational numbers. Let m be the common denominator of these fractions. We introduce a new variable t:

$$\frac{ax+b}{cx+\mathrm{d}} = t^m.$$

Evidently, after this, x, dx/dt and the expressions:

$$\left(\frac{ax+b}{cx+d}\right)^{\lambda}, \quad \left(\frac{ax+b}{cx+d}\right)^{\mu}$$

will be rational functions of t, and (6) reduces to the integral of a rational fraction.

2. *Binomial differential. The integrals of binomial differentials :*

$$\int x^m (a + bx^n)^p \, dx, \tag{7}$$

where m, n and p are rational numbers, reduce in many cases to the integral (6).

We put $x = t^{1/n}$:

$$\int x^m (a + bx^n)^p \, dx = \frac{1}{n} \int t^{\frac{m+1}{n} - 1} (a + bt)^p \, dt.$$

If p or $(m+1)/n$ is an integer, the integral obtained is of the type (6).

It follows from the evident equality:

$$\int t^{(\overline{m+1}/n)-1} (a + bt)^p \, dt = \int t^{(\overline{m+1}/n)+p-1} \left(\frac{a+bt}{t}\right)^p dt$$

that in the case when $\frac{m+1}{n} + p$ is an integer, integral (7) reduces to the form (6).

There is a *theorem of Tchebysheff*, according to which the above three cases exhaust all the cases in which the integral of a binomial differential is expressible in terms of elementary functions.

199. Integrals of the type $\int R(x, \sqrt{ax^2 + bx + c})dx$. An integral of the form

$$\int R\left(x, \sqrt{ax^2 + bx + c}\right) dx, \tag{8}$$

where R is a rational function of its arguments, leads to the integral of a rational fraction with the aid of *Euler's substitution*.

In the case $a > 0$, *Euler's first substitution* can be used:

$$\sqrt{ax^2 + bx + c} = t - \sqrt{a}x.$$

On squaring both sides of this equation and solving with respect to x, we get:

$$x = \frac{t^2 - c}{2t\sqrt{a} + b},$$

from which it is clear that x, dx/dt, and $\sqrt{ax^2 + bx + c}$ are rational functions of t, and therefore integral (8) leads to the integral of a rational fraction.

In the case $c > 0$, we can use *Euler's second substitution* :

$$\sqrt{ax^2 + bx + c} = tx + \sqrt{c}.$$

We leave it to the reader to verify this.

In the case $a < 0$, the quadratic expression $(ax^2 + bx + c)$ must have real zeros x_1 and x_2, since it would otherwise have the $(-)$ sign for all real x, whilst $\sqrt{ax^2 + bx + c}$ would be imaginary. In the case of real zeros integral (8) leads to integration of a rational fraction with the aid of *Euler*'s *third substitution* :

$$\sqrt{a(x - x_1)(x - x_2)} = t(x - x_2),$$

which we suggest may also be checked by the reader.

Euler's substitutions lead to complicated expressions for the most part, and we therefore give a second method of evaluating integral (8).

For brevity, let

$$y = \sqrt{ax^2 + bx + c}.$$

Every positive even power of y consists of a polynomial in x, and the integrand is therefore easily reduced to the form:

$$R(x, y) = \frac{\omega_1(x) + \omega_2(x)\,y}{\omega_3(x) + \omega_4(x)\,y},$$

where the $\omega_s(x)$ are polynomials in x. If we get rid of the irrational in the denominator and carry out elementary transformations, we can put the expression written in the form:

$$R(x, y) = \frac{w_5(x)}{w_6(x)} + \frac{w_7(x)}{w_8(x)\,y}.$$

The first term is a rational fraction, which we know how to integrate. On dividing out $\omega_7(x)/\omega_8(x)$ into an integral part and reducing the remaining proper fraction to partial fractions we arrive at an integral of the form:

$$\int \frac{\varphi(x)}{\sqrt{ax^2 + bx + c}}\,dx \tag{9}$$

and at another of the form

$$\int \frac{dx}{(x - a)^n \sqrt{ax^2 + bx + c}}, \tag{10}$$

where $\varphi(x)$ is a polynomial in x.

We assume here that polynomials $\omega_s(x)$ only have real zeros.

Before we consider integrals (9) and (10) in general, we note two simple particular cases of integral (9):

$$\int \frac{dx}{\sqrt{ax^2+bx+c}} =$$
$$= \frac{1}{\sqrt{a}} \log\left(x + \frac{b}{2a} + \sqrt{x^2 + \frac{b}{a}x + \frac{c}{a}}\right) + C \quad (a > 0), \tag{11}$$

$$\int \frac{dx}{\sqrt{-x^2+bx+c}} = \int \frac{dx}{\sqrt{m^2 - \left(x - \frac{b}{2}\right)^2}} = \arcsin \frac{x - \frac{1}{2}b}{m} + C. \tag{12}$$

Formula (11) is easily obtained with the aid of Euler's first substitution, whilst we worked out (12) previously in [92].

Integral (9) can conveniently be worked out by using the formula:

$$\int \frac{\varphi(x)}{\sqrt{ax^2+bx+c}}\, dx = \psi(x)\sqrt{ax^2+bx+c} + \lambda \int \frac{dx}{\sqrt{ax^2+bx+c}}, \tag{13}$$

where $\psi(x)$ is a polynomial whose degree is one less than that of $\varphi(x)$, and λ is a constant. We shall not dwell on the proof of (13). By differentiating (13) and clearing fractions, we obtain an identity between two polynomials, from which the coefficients of $\psi(x)$ and λ can be determined.

Integral (10) reduces to integral (9) with the aid of the substitution

$$x - a = \frac{1}{t}.$$

Example.

$$\int \frac{dx}{x + \sqrt{x^2 - x + 1}} = \int \frac{x - \sqrt{x^2 - x + 1}}{x - 1}\, dx =$$
$$= \int \frac{x}{x-1}\, dx - \int \frac{x^2 - x + 1}{(x-1)\sqrt{x^2 - x + 1}}\, dx =$$
$$= x + \log(x-1) - \int \frac{x^2 - x + 1}{(x-1)\sqrt{x^2 - x + 1}}\, dx.$$

But

$$\frac{x^2 - x + 1}{x - 1} = x + \frac{1}{x-1},$$

and therefore

$$\int \frac{x^2 - x + 1}{(x-1)\sqrt{x^2 - x + 1}}\, dx = \int \frac{x}{\sqrt{x^2 - x + 1}}\, dx + \int \frac{dx}{(x-1)\sqrt{x^2 - x + 1}}.$$

In accordance with (13):

$$\int \frac{x}{\sqrt{x^2 - x + 1}} \, dx = a\sqrt{x^2 - x + 1} + \lambda \int \frac{dx}{\sqrt{x^2 - x + 1}}$$

We differentiate this relationship and clear fractions, and obtain the identity:

$$2x = a(2x - 1) + 2\lambda \, ,$$

whence

$$a = 1; \quad \lambda = \frac{1}{2} ,$$

and therefore, by (11):

$$\int \frac{x}{\sqrt{x^2 - x + 1}} \, dx = \sqrt{x^2 - x + 1} + \frac{1}{2} \log \left(x - \frac{1}{2} + \sqrt{x^2 - x + 1} \right) + C.$$

On substituting

$$x - 1 = \frac{1}{t} ,$$

we obtain:

$$\int \frac{dx}{(x - 1)\sqrt{x^2 - x + 1}} = -\int \frac{dt}{\sqrt{t^2 + t + 1}} =$$

$$= -\log \left(t + \frac{1}{2} + \sqrt{t^2 + t + 1} \right) + C =$$

$$= -\log \left(\frac{1}{x - 1} + \frac{1}{2} + \sqrt{\frac{1}{(x - 1)^2} + \frac{1}{x - 1} + 1} \right) + C =$$

$$= -\log \left(x + 1 + 2\sqrt{x^2 - x + 1} \right) + \log (x - 1) + C.$$

Finally:

$$\int \frac{dx}{x + \sqrt{x^2 - x + 1}} = x - \sqrt{x^2 - x + 1} - \frac{1}{2} \log \left(x - \frac{1}{2} + \sqrt{x^2 - x + 1} \right) +$$

$$+ \log \left(x + 1 + 2\sqrt{x^2 - x + 1} \right) + C.$$

Integral (8) is a particular case of an *Abel integral*, which has the form

$$\int R(x, y) \, dx \, , \tag{14}$$

where R is a rational function of its arguments and y *is an algebraic function of* x, i.e. a function of x, defined by the equation

$$f(x, y) = 0 \, , \tag{15}$$

where the left-hand side is an integral polynomial in x and y. If

$$y = \sqrt{P(x)},$$

where $P(x)$ is a polynomial of the third or fourth degree in x, the Abel integral (14) is called an elliptic integral. We discuss these latter in the third volume.

They are not generally expressible in terms of elementary functions, and the same applies all the more for the general Abel integral. If the degree of the polynomial $P(x)$ is higher than the fourth, integral (14) is called hyper-elliptic.

If (15), expressing y as an algebraic function of x, has the property that x and y can be expressed in the form of rational functions of an auxiliary parameter t, integral (14) evidently reduces to the integral of a rational fraction. In this case, the algebraic curve corresponding to (15) is referred to as *unicursal*. In particular, Euler's substitutions serve to prove the unicursality of the curve:

$$y^2 = ax^2 + bx + c.$$

200. Integrals of the form $\int R(\sin x, \cos x)\mathrm{d}x$. Integrals of the form:

$$\int R(\sin x, \cos x)\, \mathrm{d}x\,, \tag{16}$$

where R is a rational function of its arguments, reduce to integrals of rational fractions on introducing the new variable

$$t = \tan \frac{1}{2}\, x.$$

We obtain, in fact, with the aid of well known trigonometric formulae:

$$\sin x = \frac{2t}{1+t^2}\,; \quad \cos x = \frac{1-t^2}{1+t^2}\,,$$

and in addition,

$$x = 2 \operatorname{arc\,tan} t; \quad \mathrm{d}x = \frac{2\mathrm{d}t}{1+t^2}\,,$$

from which our assertion follows at once.

We now notice some particular cases, where the working can be simplified.

1. Suppose that $R(\sin x, \cos x)$ remains unchanged on replacing $\sin x$ and $\cos x$ by $(-\sin x)$ and $(-\cos x)$ respectively, i.e. that $R(\sin x, \cos x)$ has period π.

Since

$$\sin x = \cos x \tan x\,,$$

$R(\sin x, \cos x)$ is a rational function of $\cos x$ and $\tan x$, which remains unchanged on replacing $\cos x$ by $(-\cos x)$, i.e. it contains only even powers of $\cos x$:

$$R(\sin x, \cos x) = R_1(\cos^2 x, \tan x)\,.$$

It is sufficient in the present case to substitute

$$t = \tan x\,,$$

in order to reduce (16) to the integral of a rational fraction.

With this:

$$dx = \frac{dt}{1+t^2}\,;\ \cos^2 x = \frac{1}{1+t^2}\,.$$

Thus, if $R(\sin x, \cos x)$ *remains unchanged on replacing* $\sin x$ and $\cos x$ by $(-\sin x)$ and $(-\cos x)$ *respectively, integral* (16) *reduces to the integral of a rational fraction with the aid of the substitution* $t = \tan x$.

2. We now suppose that $R(\sin x, \cos x)$ merely changes sign on replacing $\sin x$ by $(-\sin x)$. The function

$$\frac{R(\sin x, \cos x)}{\sin x}$$

will be completely unchanged with this replacement, i.e. contains only even powers of $\sin x$, so that:

$$R(\sin x, \cos x) = R_1(\sin^2 x, \cos x)\cdot \sin x\,.$$

On substituting $t = \cos x$, we get:

$$\int R(\sin x, \cos x)\,dx = -\int R_1(1 - t^2, t)\,dt\,,$$

i.e. *if* $R(\sin x, \cos x)$ *merely changes sign on replacing* $\sin x$ *by* $(-\sin x)$, *integral* (16) *reduces to the integral of a rational fraction by the substitution* $t = \cos x$.

3. It can easily be shown in the same way that if $R(\sin x, \cos x)$ *merely changes sign on replacing* $\cos x$ *by* $(-\cos x)$, (16) *reduces to the integral of a rational fraction by substituting* $t = \sin x$.

201. Integrals of the form $\int e^{ax}[P(x)\cos bx + Q(x)\sin bx]dx$. Integration by parts of an integral of the form:

$$\int e^{ax}\varphi(x)\,dx\,, \tag{17}$$

where $\varphi(x)$ is a polynomial of the nth degree in x, gives

$$\int e^{ax}\varphi(x)\,dx = \frac{1}{a}e^{ax}\varphi(x) - \frac{1}{a}\int e^{ax}\varphi'(x)\,dx.$$

Having thus separated out from the integral a term in the form of a product of e^{ax} and a polynomial of degree n, we are able to lower

the degree of the polynomial under the integral sign by unity. If we continue to integrate by parts, whilst noting that

$$\int e^{ax}\,dx = \frac{1}{a}\,e^{ax} + C,$$

we obtain:

$$\int e^{ax}\,\varphi(x)\,dx = e^{ax}\,\psi(x) + C\,, \tag{18}$$

where $\psi(x)$ is a polynomial of the same degree n as $\varphi(x)$, i.e. *a product of the exponential function* e^{ax} *and a polynomial of degree* n *retains the same form on integration.*

On differentiating (18) and cancelling e^{ax} from both sides of the identity obtained, we can find the coefficients of the polynomial $\psi(x)$ by the method of undetermined coefficients.

We now consider the integral of more general type:

$$\int e^{ax}[P(x)\cos bx + Q(x)\sin bx]\,dx\,, \tag{19}$$

where $P(x)$ and $Q(x)$ are polynomials in x. Let n be the greater of the degrees of these polynomials. If we use the expedient of introducing complex numbers, we can reduce integral (19) to integral (17) by substituting for $\cos bx$ and $\sin bx$ in accordance with Euler's formula [176]:

$$\cos bx = \frac{e^{bxi} + e^{-bxi}}{2}\,;\ \sin bx = \frac{e^{bxi} - e^{-bxi}}{2i}\,.$$

This gives:

$$\int e^{ax}[P(x)\cos bx + Q(x)\sin bx]\,dx =$$
$$= \int e^{(a+bi)x}\,\varphi(x)\,dx + \int e^{(a-bi)x}\,\varphi_1(x)\,dx,$$

where $\varphi(x)$ and $\varphi_1(x)$ are polynomials of degrees not higher than n. Application of formula (18) gives:

$$\int e^{ax}[P(x)\cos bx + Q(x)\sin bx]\,dx = e^{(a+bi)x}\,\psi(x) + e^{(a-bi)x}\,\psi_1(x) + C\,,$$

where $\psi(x)$ and $\psi_1(x)$ are polynomials of degrees not higher than n. On substituting:

$$e^{\pm bxi} = \cos bx \pm i\sin bx\,,$$

we finally get:

$$\int e^{ax}[P(x)\cos bx + Q(x)\sin bx]\,dx =$$
$$= e^{ax}[R(x)\cos bx + S(x)\sin bx] + C\,, \tag{20}$$

where $R(x)$ and $S(x)$ are polynomials of degrees not higher than n. Hence we see that *the integrand in* (19) *retains the same form after integration, with polynomials of degrees equal to the higher of the degrees of the original polynomials.*

We differentiate (20), cancel e^{ax} in the identity obtained, and equate coefficients of like terms of the form $x^s \cos bx$ and $x^s \sin bx$ ($s = 0, 1, 2, \ldots, n$) on each side; we thus get a system of linear equations defining the coefficients of the polynomials $R(x)$ and $S(x)$. We remark here that, if either $\cos bx$ or $\sin bx$ is absent from the integrand, both trigonometric functions still have to be written on the right-hand side of the formula, bearing in mind the rule given above concerning the degrees of polynomials $R(x)$ and $S(x)$.

Integrals of the form:

$$\int e^{ax}\varphi(x)\sin(a_1 x + b_1)\sin(a_2 x + b_2)\ldots$$
$$\ldots\cos(c_1 x + d_1)\cos(c_2 x + d_2)\ldots dx$$

reduce at once to integrals of type (19).

In fact, by using the familiar trigonometric formulae that express the sums and differences of sines and cosines as products, we can conversely express the product of any two of the above trigonometric functions as a sum or difference of sines or cosines. Repeated application of this transformation enables us to reduce the number of trigonometric factors under the integral sign to one, and hence we obtain an integral of type (19).

Example. By (20):

$$\int e^{ax}\sin bx\,dx = e^{ax}(A\cos bx + B\sin bx) + C.$$

We differentiate and cancel e^{ax}:

$$\sin bx = (aA + bB)\cos bx + (-bA + aB)\sin bx\,,$$

whence

$$aA + bB = 0\,,\qquad -bA + aB = 1\,,$$

i.e.

$$A = -\frac{b}{a^2+b^2}\,;\quad B = \frac{a}{a^2+b^2}\,,$$

and finally:

$$\int e^{ax}\sin bx\,dx = e^{ax}\left(-\frac{b}{a^2+b^2}\cos bx + \frac{a}{a^2+b^2}\sin bx\right) + C. \qquad (21)$$

ANSWERS

Chapter I

3. (*a*) $-2 < x < 4$; (*b*) $-1 < x < 0$; (*c*) $x < -3, x > 1$; (*d*) $x > 0$.
4. -24; -6; 0; 0; 0; 6.
5. $1; 5/4; \sqrt{1+x^2}; |x|^{-1}\sqrt{1+x^2}; 1/\sqrt{1+x^2}$.
6. $\pi; \frac{1}{2}\pi; 0$.
7. $f(x) = \frac{1}{3}(1-5x)$.
8. $f(x) = \frac{7}{6}x^2 - \frac{13}{6}x + 1$.
9. 0.4.
10. $\frac{1}{2}(x+|x|)$.
11. (*a*) $-1 \leqslant x < +\infty$; (*b*) $-\infty < x < +\infty$.
12. $(-\infty, -2)$, $(-2, 2)$, $(2, \infty)$ i.e. everywhere except $x = \pm 2$.
13. (*a*) $-\infty < x \leqslant -\sqrt{2}$, $\sqrt{2} \leqslant x < \infty$; (*b*) $x = 0$, $|x| \geqslant \sqrt{2}$.
14. $-1 \leqslant x \leqslant 2$.
15. $-2 < x \leqslant 0$.
16. $-\infty < x \leqslant -1$, $0 \leqslant x \leqslant 1$.
17. $-2 < x < 2$.
18. $-1 < x < 1$, $2 < x < +\infty$.
19. $-\frac{1}{3} \leqslant x \leqslant 1$.
20. $1 \leqslant x \leqslant 100$.
21. $k\pi \leqslant x \leqslant \left(k + \frac{1}{2}\right)\pi$, $(k = 0, \pm 1, \pm 2, \ldots)$.
22. $\Phi(x) = 2x^4 - 5x^2 - 10$; $\psi(x) = -3x^3 + 6x$.
23. $\Phi[\psi(x)] = 2^{2x}$; $\psi[\Phi(x)] = 2^{x^2}$.
24. x.
25. $(x+2)^2$.
30. $-\pi/2$; 0; $\pi/4$.
31. (*a*) $y = 0$ if $x = -1$, $y > 0$ if $x > -1$, $y < 0$ if $x < -1$; (*b*) $y = 0$ if $x = 2$ and if $x = -1$; $y > 0$ if $-1 < x < 2$, $y < 0$ if $-\infty < x < -1$ and if $2 < x < +\infty$; (*c*) $y > 0$ if $-\infty < x < \infty$; (*d*) $y = 0$ if $x = 0$, $-\sqrt{3}$ or $\sqrt{3}$, $y > 0$ if $-\sqrt{3} < x < 0$ and $\sqrt{3} < x < +\infty$; $y < 0$ if $-\infty < x < -\sqrt{3}$ and if $0 < x < \sqrt{3}$; (*e*) $y = 0$ if $x = 1$; $y > 0$ if $-\infty < x < -1$ and if $1 < x < \infty$, $y < 0$ if $0 < x < 1$.

32. (*a*) $x = \frac{1}{2}(y-3)$ $(-\infty < y < \infty)$; (*b*) $x = \sqrt{y+1}$ and $x = -\sqrt{(y+1)}$ $(-1 \leqslant y \leqslant +\infty)$; (*c*) $x = \sqrt[3]{1-y^3}$, $(-\infty < y < \infty)$; (*d*) $x = 2 \cdot 10^y$, $(-\infty < y < +\infty)$; (*e*) $x = \frac{1}{3}\tan y$ $\left(-\frac{1}{2}\pi < y < \frac{1}{2}\pi\right)$.

33. $x = y$ if $-\infty < y \leqslant \infty$; $x = \sqrt{y}$ if $0 < y < +\infty$.

34. (*a*) $y = u^{10}$, $u = 2x - 5$; (*b*) $y = 2^u$, $u = \cos x$; (*c*) $y = \log_{10} u$, $u = \tan v$, $v = \frac{1}{2}x$; (*d*) $y = \arcsin u$, $u = 3^v$, $v = -x^2$.

35. (*a*) $y = \sin^2 x$; (*b*) $y = \arctan\sqrt{\log_{10} x}$; (*c*) $y = 2(x^2-1)$ if $|x| \leqslant 1$ and $y = 0$ if $|x| > 1$.

36. (*a*) $y = -\cos x^2$, $\sqrt{\pi} \leqslant x \leqslant \sqrt{2\pi}$; (*b*) $y = \log_{10}(10 - 10^x)$, $-\infty < < x < 1$; (*c*) $y = \frac{1}{3}x$ if $-\infty < x < 0$ and $y = x$ if $0 \leqslant x < +\infty$.

81. $n > 1/\sqrt{\varepsilon}$. (*a*) $n \geqslant 4$; (*b*) $n > 10$; (*c*) $n \geqslant 32$.

82. $n > 1/\varepsilon - 1 = N$. (*a*) $N = 9$; (*b*) $N = 99$; (*c*) $N = 999$.

83. $\delta = \varepsilon/5$ $(\varepsilon < 1)$. (*a*) 0.02; (*b*) 0.002; (*c*) 0.0002.

84. (*a*) $\log x < -N$ for $0 < x < \delta(N)$;
(*b*) $2^x > N$ for $x > X(N)$;
(*c*) $|f(x)| > N$ for $|x| > X(N)$.

85. (*a*) 0; (*b*) 1; (*c*) 2; (*d*) 7/30.

86. 1/2. **87.** 1. **88.** $-3/2$. **89.** 1. **90.** 3. **91.** 1. **92.** 3/4. **93.** 1/3.
94. 0. **95.** 0. **96.** 1. **97.** 0. **98.** ∞. **99.** 0. **100.** 1.
101. 1/2. **102.** 3. **103.** 4/3. **104.** 1/9. **105.** $-1/56$. **106.** 12. **107.** 3/2.
108. $-1/3$
109. 1. **110.** $\frac{1}{2\sqrt{x}}$. **111.** $\frac{1}{3\sqrt{x^2}}$ **112.** 0. **113.** $\frac{1}{2}a$. **114.** $-5/2$. **115.** 1/2.
116. 0. **117.** (*a*) $\frac{1}{2}\sin 2$; (*b*) 0. **118.** 3. **119.** 5/2. **120.** 1/3. **121.** π. **122.** 1/2.
123. $\cos a$. **124.** $-\sin a$. **125.** 0. **126.** $-1/\sqrt{3}$. **127.** 1. **128.** e^{-1}. **129.** e^2.
130. e^{-1}. **131.** e^{-4}. **132.** e^x. **133.** e. **140.** (*a*) $x \neq \frac{1}{2}\pi + k\pi$, where k is an integer; (*b*) $x \neq k\pi$. **143.** $A = 4$. **144.** (*a*) $f(0) = n$; (*b*) $f(0) = 1/2$; (*c*) $f(0) = 2$; (*d*) $f(0) = 2$; (*e*) $f(0) = 0$; (*f*) $f(0) = 1$.

Chapter II

1. (*a*) 3; (*b*) 0.21; (*c*) $2h + h^2$. **2.** (*a*) 0.1; (*b*) -3; (*c*) $\sqrt[3]{a+h} - \sqrt[3]{a}$.
4. (*a*) 624; 1560; (*b*) 0.01; 100; (*c*) -1; 0.00011.

5. (*a*) $a\,\Delta x$; a; (*b*) $3x^2\,\Delta x + 3x(\Delta x)^2 + (\Delta x)^3$; $3x^2 + 3x\,\Delta x + (\Delta x)^2$;

(*c*) $-\dfrac{2x\,\Delta x + (\Delta x)^2}{x^2\,(x+\Delta x)^2} = -\dfrac{2x+\Delta x}{x^2\,(x+\Delta x)^2}$; (*d*) $\sqrt{(x+\Delta x)} - \sqrt{x}$;

$\dfrac{1}{\sqrt{x}+\sqrt{(x+\Delta x)}}$

(*e*) $2^x\,(2^{\Delta x} - 1)$; $2^x\,\dfrac{2^{\Delta x}-1}{\Delta x}$; (*f*) $\log\dfrac{x+\Delta x}{x}$; $\dfrac{1}{\Delta x}\log\dfrac{x+\Delta x}{x}$

6. (*a*) -1; (*b*) 0.1; (*c*) h; 0. **7.** 15 cm/sec.

8. $7\cdot 5$. **9.** $\dfrac{f(x+\Delta x)-f(x)}{\Delta x}$. **10.** $\lim\limits_{\Delta x\to 0}\dfrac{f(x+\Delta x)-f(x)}{\Delta x} = f'(x)$ is given.

11. (*a*) $\Delta\varphi/\Delta t$; (*b*) $\mathrm{d}\Phi/\mathrm{d}t = \lim\limits_{\Delta t\to 0}\dfrac{\Delta\Phi}{\Delta t}$ where Φ is the magnitude of the angle of rotation at the instant t.

12. (*a*) $\Delta T/\Delta t$, (*b*) $\mathrm{d}T/\mathrm{d}t = \lim\limits_{\Delta t\to 0}\Delta T/\Delta t$ where T is the temperature at the time t. **13.** $\mathrm{d}Q/\mathrm{d}t = \lim\limits_{\Delta t\to 0}\Delta Q/\Delta t$ where Q is the quantity of matter at the time t.

14. (*a*) $\Delta m/\Delta x$, (*b*) $\lim\limits_{x\to 0}\Delta m/\Delta x$. **15.** (*a*) $-1/6$; (*b*) $-5/21$; (*c*) $50/201$; $y'(2) = -1/4$.

16. $\sec^2 x$. **17.** (*a*) $3x^2$; (*b*) $-2/x^3$; (*c*) $\dfrac{1}{2\sqrt{x}}$; (*d*) $-\operatorname{cosec}^2 x$.

18. 1/12. **19.** $f'(0) = -8$, $f'(1) = 0$, $f'(2) = 0$.

20. 30 m/sec. **21.** -1; 2; $\tan\Phi = 3$. **22.** (*a*) $\pm\infty$; (*b*) ∞; (*c*) -1: $f'(k\pi + \frac{1}{2}\pi) = -1$, $f'(k\pi + \frac{1}{2}\pi) = 1$.

23. $5x^4 - 12x^2 + 2$. **24.** $-\dfrac{1}{3} + 2x - 2x^2$. **25.** $2ax + b$.

26. $-15x^2/a$. **27.** $mat^{m-1} + (m+n)\,bt^{m+n-1}$. **28.** $6ax^5$. **29.** $-\pi/x^2$.

30. $2x^{-1/3} - 5x^{+3/2} - 3x^{-4}$.

31. $\dfrac{8}{3}x^3\sqrt{x^2}$. **32.** $\dfrac{4}{3}bx^{-7/3} - \dfrac{2}{3}ax^{-5/3}$.

33. $(bc - ad)/(c + dx)^2$. **34.** $-(2x^2 + 6x - 25)/(x^2 - 5x + 5)^2$.

35. $(1 - 4x)/x^2(2x - 1)^2$. **36.** $\dfrac{1}{\sqrt{z}\,(1-\sqrt{z})^2}$.

37. $5\cos x - 3\sin x$. **38.** $4\operatorname{cosec}^2 2x$. **39.** $-2(\sin x - \cos x)^{-2}$.

40. $t^2\sin t$. **41.** 0. **42.** $\cot x - x\operatorname{cosec}^2 x$. **43.** $\arcsin x + x/\sqrt{1-x^2}$.

44. $x\arctan x$. **45.** $(x + 7)\,x^6\,e^x$. **46.** xe^x. **47.** $(x - 2)\,x^{-3}\,e^x$.

48. $(5x^4 - x^5)\,e^{-x}$. **49.** $e^x(\cos x - \sin x)$. **50.** $x^2\,e^x$.

51. $e^x\,(\arcsin x + 1/\sqrt{1-x^2})$ **52.** $x(2\log x - 1)/\log^2 x$.

53. $3x^2\log x$. **54.** $y = 2/x + \log x/x^2 - 2/x^2$.

55. $3ac^{-3}(ax + b)^2$. **56.** $12ab + 18b^2\,y$. **57.** $16x(3 + 2x^2)^3$.

58. $\frac{x^2-1}{(2x-1)^8}$. **59.** $-\frac{x}{\sqrt{1-x^2}}$. **60.** $\frac{bx^2}{\sqrt[3]{(a+bx^3)^2}}$

61. $-\sqrt{(a/x)^{2/3}-1}$

62. $(1-\tan^2 x+\tan^4 x)\sec^2 x$. **63.** $-\frac{1}{2}\operatorname{cosec}^2 x\,(\cot x)^{-1/2}$.

64. $2-15\cos^2 x\sin x$. **65.** $-16\cos 2t\operatorname{cosec}^3 2t$. **66.** $\sin x/(1-3\cos x)^3$

67. $\sin^3 x/\cos^4 x$. **68.** $\frac{1}{2}(3\cos x+2\sin x)/\sqrt{15\sin x-10\cos x}$.

69. $\frac{2}{3}\cos x\sin^{-1/3}x+3\sin x\cos^{-4}x$.

70. $1/2\sqrt{(1-x^2)(1+\arcsin x)}$. **71.** $\frac{1}{2(1+x^2)\sqrt{\arctan x}}-\frac{3(\arcsin x)}{\sqrt{1-x^2}}$.

72. $-\frac{1}{(1+x^2)(\arctan x)^2}$. **73.** $\frac{xe^x+e^x+1}{2\sqrt{xe^x+x}}$. **74.** $\frac{2e^x-2^x\log 2}{3\sqrt[3]{(2e^x-2^x+1)^2}}+$
$+\frac{5\log^4 x}{x}$.

75. $15\sin^2 5x\cos 5x\cos^2 x/3-\frac{2}{3}\sin^3 5x\cos x/3\sin x/3$.

76. $\frac{4x+3}{(x-2)^3}$. **77.** $\frac{x^2+4x-6}{(x-3)^5}$.

78. $x^7(1-x^2)^{-5}$. **79.** $\frac{x-1}{x^2\sqrt{2x^2-2x+1}}$. **80.** $(a^2+x^2)^{-3/2}$.

81. $x^2(1+x^2)^{-5/2}$.

82. $x^{-1/3}(1+\sqrt{x})^3$. **83.** $x^5(1+x^3)^{2/3}$. **84.** $(x-1)^{-3/4}(x+2)^{-5/4}$.

85. $4x^3(a-2x^3)(a-5x^3)$.

86. $\frac{2abmnx^{n-1}(a+bx^n)^{m-1}}{(a-bx^n)^{m+1}}$. **87.** $\frac{x^3-1}{(x+2)^6}$. **88.** $\frac{a-3x}{2\sqrt{a-x}}$.

89. $\frac{3x^2+2(a+b+c)x+ab+bc+ca}{2\sqrt{[(x+a)(x+b)(x+c)]}}$. **90.** $\frac{1+2\sqrt{y}}{6\sqrt{y}(y+\sqrt{y})^{2/3}}$.

91. $2(7t+4)(3t+2)^{1/3}$. **92.** $(y-a)(2ay-y^2)^{-3/2}$. **93.** $(1+e^x)^{-1/2}$.

94. $\sin^3 x\cos^2 x$. **95.** $(\sin x\cos x)^{-4}$. **96.** $10\tan 5x\sec^2 5x$.

97. $x\cos x^2$. **98.** $3t^2\sin 2t^3$. **99.** $3\cos x\cos 2x$. **100.** $\tan^4 x$.

101. $\frac{\cos 2x}{\sin^4 x}$. **102.** $\frac{(\alpha-\beta)\sin 2x}{2\sqrt{[\alpha\sin^2 x+\beta\cos^2 x]}}$.

103. 0. **104.** $\frac{1}{2}(1-x^2)^{-1/2}\arcsin(2\arccos x-\arcsin x)$.

105. $2x^{-1}(2x^2-1)^{-1/2}$. **106.** $(1+x^2)^{-1}$.

107. $(1-x^2)^{-3/2}[x\arccos x-\sqrt{1-x^2}]$. **108.** $(a-bx^2)^{-1/2}$.

109. $\sqrt{(a-x)/(a+x)}$. **110.** $2\sqrt{a^2-x^2}$. **111** $-x(2x-x^2)^{-1/2}$.

112. $\arcsin\sqrt{x}$.

113. $\frac{5}{\sqrt{(1-25x^2)}\arcsin 5x}$. **114.** $\frac{1}{x\sqrt{(1-\log^2 x)}}$. **115.** $\frac{\sin\alpha}{1-2x\cos\alpha+x^2}$.

116. $(5+4\sin x)^{-1}$. **117.** $4x\sqrt{[x/(b-x)]}$. **118.** $\sin^2 x/(1+\cos^2 x)$.

119. $\frac{1}{2}a\,e^{\frac{1}{2}ax}$. **120.** $\sin 2x e^{\sin^2 x}$. **121.** $2m^2p(2ma^{mx}+b)^{p-1}a^{mx}\log a$.

122. $(a \cos \beta t - \beta \sin \beta t)\, e^{at}$.

123. $e^{ax} \sin \beta x$. **124.** $e^{-x} \cos 3x$. **125.** $x^{n-1}\, a^{-x^2}(n - 2x^2 \log a)$.

126. $-\frac{1}{2} \sin x \cos^{-1/2} x\, a^{\sqrt{\cos x}} (1 + \log a \sqrt{\cos x})$.

127. $3x^{-2} \left(\sin \frac{1}{x}\right)^{-2} \cot \frac{1}{x} \log 3$.

128. $(2ax + b)/(ax^2 + bx + c)$. **129.** $(a^2 + x^2)^{-1/2}$. **130.** $\sqrt{x}/(1 + \sqrt{x})$.

131. $(x^2 + 2ax)^{-1/2}$. **132.** $-2x^{-1} \log^{-3} x$. **133.** $-x^2 \tan[(x - 1)/x]$.

134. $(2x + 11)/(x^2 - x - 2)$. **135.** $\dfrac{3x^2 - 16x + 19}{(x-1)(x-2)(x-3)}$.

136. $1/(\sin^3 x \cos x)$. **137.** $\sqrt{x^2 - a^2}$.

138. $\dfrac{-6x^2}{(3 - 2x^3) \log (3 - 2x^3)}$ **139.** $\dfrac{15a \log^2 (ax+b)}{(ax + b)}$.

140. $2(x^2 + a^2)^{-1/2}$. **141.** $(mx + n)/(x^2 - a^2)$. **142.** $\sqrt{2} \sin \log x$.

143. $\sin^{-3} x$. **144.** $x^{-1} \sqrt{1 + x^2}$. **145.** $(x + 1)/(x^3 - 1)$.

146. $3/\sqrt{1 - 9x^2}\, [2^{\arcsin 3x} \log 2 + 2(1 - \arccos 3x)]$.

147. $(3^{\sin ax/\cos bx} \log 3 + \sin^2 ax/\cos^2 bx)\, x\, \dfrac{a \cos ax \cos bx + b \sin ax \sin bx}{\cos^2 bx}$.

148. $(1 + 2 \sin x)^{-1}$. **149.** $x^{-1}(1 + \log^2 x)^{-1}$.

150. $\dfrac{1}{\sqrt{1 - x^2} \cdot \arcsin x} + \dfrac{\log x}{x} + \dfrac{1}{x\sqrt{1 - \log^2 x}}$. **151.** $\dfrac{-1}{x(1 + \log^2 x)}$.

152. $x^2/(x^4 + x^2 - 2)$. **153.** $2 \cos^{-1} x \sin^{-1/2} x$. **154.** $(x^2 - 3x)/(x^4 - 1)$.

155. $(1 + x^3)^{-1}$. **156.** $(1 - x^2)^{-3/2} \arcsin x$.

157. $y' = 1$ if $x > 0$, $y' = -1$ if $x < 0$; $y'(0)$ does not exist. **158.** $|2x|$.

159. $1/x$. **160.** $f'(x) = -1$, $x \leqslant 0$, $f'(x) = -e^{-x}$, $x > 0$.

161. $1 - x$. **162.** $2 + \frac{1}{4}(x - 3)$. **163.** -1. **164.** 0.

170. $(1 + 2x)(1 + 3x) + 2(1 + x)(1 + 3x) + 3(x + 1)(1 + 2x)$.

171. $-(x + 2)(5x^2 + 19x + 20)(x + 1)^{-4}(x + 3)^{-5}$.

172. $\frac{1}{2}(x^2 - 4x + 2)[x(x - 1)(x - 2)^3]^{1/2}$.

173. $\frac{1}{3}(3x^2 + 5)x^{2/3}(x^2 + 1)^{-4/3}$. **174.** $\sqrt[x]{x}\, \dfrac{1 - \log x}{x^2}$.

175. $x^{\sqrt{x} - \frac{1}{2}} \left(1 + \frac{1}{2} \log x\right)$. **176.** $x^{x^x} x^x \left(\frac{1}{x} + \log x + \log^2 x\right)$.

177. $x^{\sin x} \left(\dfrac{\sin x}{x} + \cos x \log x\right)$.

178. $(\cos x)^{\sin x} (\cos x \log \cos x - \sin x \cdot \tan x)$.

179. $\left(1 + \frac{1}{x}\right)^x \left\{\log\left(1 + \frac{1}{x}\right) + \dfrac{1}{1 + x}\right\}$. **180.** $\frac{3}{2} t^2$. **181.** $\dfrac{t - 1}{t + 1}$.

182. $\dfrac{2t}{t^2 - 1}$. **183.** $\dfrac{t(2 - t^3)}{1 - 2t^3}$. **184.** $\frac{2}{3} t^{-1/6}$. **185.** $\dfrac{t + 1}{t(t^2 + 1)}$.

186. $\tan t$. **187.** $-b/a$. **188.** $-(b/a) \tan t$.

189. $-\tan(3t)$. **190.** $y' = -1$ if $t < 0$, $y' = +1$ if $t > 0$. **191.** $-2e^{3t}$.

192. $\tan t$. **193.** 1. **194.** ∞. **195.** No. **197.** Yes. **198.** 2/5. **199.** $-b^2\, x/a^2\, y$.
200. $-x^2/y^2$. **201.** $-x(3x+2y)/(x^2+2y)$. **202.** $-\sqrt{y/x}$. **203.** $-(y/x)^{1/3}$.
204. $(1-y^3)/(1+3xy^2+4y^3)$. **205.** $10/(10-3\cos y)$. **206.** -1.
207. $y\cos^2 y/(1-x\cos^2 y)$. **208.** $y(1-x^2-y^2)/x(1+x^2+y^2)$.
209. $(x+y)^2$. **210.** $(x+y-1)^{-1}$. **211.** $e^{y/x}+y/x$.
212. $y/(x-y)$. **213.** $(x+y)/(x-y)$.
214. $[cy+x\sqrt{x^2+y^2}]/[cx-y\sqrt{x^2+y^2}]$.
215. $y(x\log - y)/x(y\log x - x)$. **216.** $45°$; arc tan $2 \sim 63°26'$. **217.** $45°$.
218. arc tan$(2/e) \sim 36°21'$. **219.** (0, 20); (1, 15); (-2, -12).
220. $y=x^2-x+1$. **221.** $-1/11$. **222.** $(1/8, -1/16)$.
223. At (4, 2) tangent is $x-4y+4=0$, normal is $4x+y=18$; at $(4, -2)$ tangent is $x+4y+4=0$, normal is $4x-y=18$.
224. $y-5=0$; $x+2=0$. **225.** $x=1$, $y=0$.
226. $7x-10y+6$, $10x+7y-34=0$.
227. $y=0$, $(\pi+4)\,x+(\pi-4)\,y-\frac{1}{4}\sqrt{2}\,\pi^2=0$.
228. $5x+6y-13=0$; $6x-5y+21=0$. **229.** $x+y=2$.
230. At (1, 0) $y=2x-2$, $x+2y=1$; at (2, 0) $x+y=2$, $x-y=2$; at (3, 0) $y=2x-6$, $2y+x=3$.
231. $14x-13y+12=0$, $13x+14y=41$.
235. $40°36'$. **236.** At (0,0) the curves have common tangent; at (1,1) they intersect at an angle arc tan $1/7 \sim 8°8'$.
238. The subtangent and the subnormal are 2, the tangent and normal $2\sqrt{2}$.
239. $1/(\log 2)$. **242.** 3 cm/sec.; 0; -9 cm/sec. **243.** 15 cm/sec.
244. B moves towards 0 with velocity 1.5 cm/sec.
245. (*a*) $y=x\tan a-\frac{1}{2}g\,(x/v_0\cos a)^2$; (*b*) $v_0^2\sin 2a/g$;
(*c*) $\sqrt{v_0^2-2v_0\,gt\sin a+g^2\,t^2}$ making an angle $\tan^{-1}[(v_0\sin a-gt)/v_0\cos a]$ with the x-axis.
246. -0.4 units per second. **247.** (9/8, 9/2).
248. (*a*) $1.2\sqrt{10}\sim 3.8$ cm/sec; (*b* 40) cm²/sec. **249.** $\frac{1}{3}\pi$ cm/sec.
250. $56x^6+210x^4$. **251.** $2(2x^2+1)\,e^{x^2}$. **252.** $2\cos 2x$.
253. $2(1-x^2)/3(1+x^2)^2$. **254.** $-x(a^2+x^2)^{-3/2}$. **255.** 2 arc tan $x+$ $+\,2x/(1+x^2)$. **256.** $2/(1-x^2)+2x(1-x^2)^{-3/2}$ arc sin x.
261. $y'''=6$. **262.** 4320. **263.** $24(x+1)^{-5}$. **264.** $-64\sin 2x$.
266. 0; 1; 2; 2. **267.** Speeds are 5; 4.997; 4.7; accelerations are 0; -0.006; -0.6. **268.** $n!\,a^n$.
269. (*a*) $n!\,(1-x)^{-(n+1)}$; (*b*) $(-1)^{n+1}\ 1.3.5\ldots(2n-3)\ 2^{-n}\,x^{\frac{1}{2}-n}$.

270. (*a*) $\sin(x + \frac{1}{2} n\pi)$; (*b*) $2^n \cos(2x + \frac{1}{2} n\pi)$; (*c*) $(-3)^n e^{-3x}$; (*d*) $(-1)^{n-1} (n-1)!\,(1+x)^{-n}$; (*e*) $(-1)^{n+1} n!\,(1+x)^{-n-1}$; (*f*) $2n!\,(1-x)^{-n-1}$; (*g*) $2^{n-1} \sin[2x + \frac{1}{2}(n-1)\pi]$; (*h*) $(-1)^{n-1} (n-1)!\,a^n(ax+b)^{-n}$.
271. (*a*) $xe^x + ne^x$; (*b*) $2^{n-1} e^{-2x}[2(-1)^n x^2 + 2n(-1)^{n-1} x + \frac{1}{2} n(n-1)(-1)^{n-2}]$; (*c*) $(1-x^2)\cos(x + \frac{1}{2} n\pi) - 2nx \sin(x + \frac{1}{2}(n-1)\pi)$; (*d*) $(-1)^{n-1}\, 1.3.5.\ldots(2n-3)\, 2^{-n} x^{-n-\frac{1}{2}} [x - (2n-1)]$; (*e*) $(-1)^n 6(n-4)!\,x^{3-n}$ if $n \geqslant 4$.
272. $(n-1)!$ **273.** (*a*) $9t^3$; (*b*) $2(1+t^2)$; (*c*) $-\sqrt{(1-t^2)}$.
274. (*a*) $-a^{-1}\sin^{-3} t$; (*b*) $(3a\cos^4 t \sin t)^{-1}$; (*c*) $-(4a \sin^4 \frac{1}{2} t)^{-1}$; (*d*) $-(at \sin^3 t)^{-1}$.
275. (*a*) 0; (*b*) $2e^{3at}$. **276.** (*a*) $(1+t^2)(1+3t^2)$; (*b*) $t(1+t)(1-t)^{-3}$.
277. $-2e^{-t}(\cos t + \sin t)^{-3}$. **278.** 1. **280.** $\Delta y = 0.009001$; $dy = 0.009$
281. 1. **282.** $dS = 2x\,\Delta x$, $\Delta S = 2x\,\Delta x + (\Delta x)^2$.
285. At $x = 0$. **286.** No! **287.** $-\pi/72 \sim -0.0436$.
289. (*a*) 0.485; (*b*) 0.965; (*c*) 1.2; (*d*) -0.045; (*e*) 0.81 radians.
290. 565 cm^3. **291.** $\sqrt{5} \sim 2.25$; $\sqrt{17} \sim 4.13$; $\sqrt{70} \sim 8.38$; $\sqrt{640} \sim 25.3$.
292. $\sqrt[3]{10} \sim 2.16$; $\sqrt[3]{70} \sim 4.13$; $\sqrt[3]{200} \sim 5.85$.
293. (*a*) 5; (*b*) 1.1; (*c*) 0.93; (*d*) 0.9. **294.** 1.0019.
296. No; since $f'(2)$ does not exist. **297.** No; since $f(\pi/2)$ does not exist.
300. $-1/3$. **301.** ∞. **302.** 3. **303.** 1/2. **304.** $2/\pi$. **305.** 1. **306.** 0.
307. ∞, if $n > 1$; a if $n = 1$; 0 if $n < 1$. **308.** 0. **309.** 1. **310.** 1/5.
311. (*a*) 0; (*b*) 0; (*c*) 2; (*d*) e^k; (*e*), (*f*) the limit does not exist.
312. The function is continuous.
313. (*a*) $x = 0$, $y = 0$; (*b*) all points of the line $y = x$; (*c*) all points on the circumference of the circle $x^2 + y^2 = 1$; (*d*) all points on the coordinate axes.
314. $\partial z/\partial x = 3(x^2 - ay)$, $\partial z/\partial y = 3(y^2 - ax)$.
315. $\partial z/\partial x = 2y(x+y)^{-2}$, $\partial z/\partial y = -2x(x+y)^2$.
316. $\partial z/\partial x = -y/x^2$, $\partial z/\partial y = 1/x$.
317. $\partial z/\partial x = x(x^2 - y^2)^{-1/2}$, $\partial z/\partial y = -y(x^2 - y^2)^{-1/2}$.
318. $\partial z/\partial x = y^2(x^2 + y^2)^{-3/2}$, $\partial z/\partial y = -xy(x^2 + y^2)^{-3/2}$.
319. $\partial z/\partial x = -(x^2 + y^2)^{-1/2}$, $\partial z/\partial y = y(x^2 + y^2)^{-1/2} [x + \sqrt{(x^2 + y^2)}]^{-1}$.
320. $\partial z/\partial x = -y/(x^2 + y^2)$, $\partial z/\partial y = x/(x^2 + y^2)$.
321. $\partial z/\partial x = yx^{y-1}$, $\partial z/\partial y = x^y \log x$.
322. $\partial z/\partial x = -yx^{-2} z$, $\partial z/\partial y = x^{-1} z$.

323. $\partial z/\partial x = xy^2(2x^2 - 2y^2)^{1/2}/[|\,y\,|\,(x^4 - y^4)]$,
$\partial z/\partial y = -yx^2(2x^2 - 2y^2)^{1/2}/[|\,y\,|\,(x^4 - y^4)]$.
324. $\partial z/\partial x = y^{-1/2} \cot[(x + a)\, y^{-1/2}]$,
$\partial z/\partial y = -\frac{1}{2}(x + a)\, y^{-3/2} \cot[(x + a)\, y^{-1/2}]$.
325. $u_x = y^2 x^{y^2-1}$, $u_y = x^{y^2} \cdot \log x \cdot 2y$.
326. $u_x = y/(1 + xy)$, $u_y = x/(1 + xy)$.
327. $f_x = 1/2$, $f_y = 0$. **328.** $f_x = 1$, $f_y = 1/2$.
329. $\Delta f = 4\Delta x + \Delta y + 2\Delta x^2 + 2\Delta x\, \Delta y + \Delta x^2\, \Delta y$; $\mathrm{d}f = 4\mathrm{d}x + \mathrm{d}y$
(a) $\Delta f - \mathrm{d}f = 8$; (b) $\Delta f - \mathrm{d}f = 0.062$.
331. $\mathrm{d}z = 3(x^2 - y)\,\mathrm{d}x + 3(y^2 - x)\,\mathrm{d}y$.
332. $\mathrm{d}z = 2xy^3\,\mathrm{d}x + 3x^2 y^2\,\mathrm{d}y$.
333. $\mathrm{d}z = 4(x^2 + y^2)^{-2}(xy^2\,\mathrm{d}x - x^2 y\,\mathrm{d}y)$.
334. $\mathrm{d}z = \sin 2x\,\mathrm{d}x - \sin 2y\,\mathrm{d}y$.
335. $\mathrm{d}z = y^2 x^{y-1}\,\mathrm{d}x + x^y(1 + y \log x)\,\mathrm{d}y$.
336. $\mathrm{d}z = 2(x^2 + y^2)^{-1}(x\mathrm{d}x + y\mathrm{d}y)$. **337.** $\mathrm{d}f = (x + y)^{-1}(\mathrm{d}x - x\mathrm{d}y/y)$.
338. $\mathrm{d}l = 0.062$ cm, $\Delta l = 0.065$ cm.
339. (a) 1.00; (b) 4.998; (c) 0.273.
340. $\pi(\alpha g - \beta l)/g\sqrt{lg}$. **341.** $\mathrm{e}^t(t \log t - 1)/(t \log^2 t)$.
342. $y^{-1/2}\, t(6 - x/2y^2) \cot(y^{-1/2}\, x)$.
343. $2t(\log t)(\tan t) + (t + t^{-1}) \tan t + \sec^2 t(1 + t^2) \log t$.
344. 0. **345.** $(\sin x)^{\cos x}(\cos x \cot x - \sin x \log \sin x)$.

Chapter III

1. $\frac{5}{7} a^2 x^2 + c$. **2.** $2x^3 + 4x^2 + 3x + c$.
3. $\frac{1}{4} x^4 + \frac{1}{3}(a + b)\, x^3 + \frac{1}{2} abx^2 + c$.
4. $a^2 x + \frac{1}{2} abx^4 + \frac{1}{7} b^2 x^7 + c$. **5.** $2x\sqrt{(2px)}/3 + c$.
6. $\frac{n}{n-1} x^{\frac{n-1}{n}} + c$. **7.** $(nx)^{\frac{1}{n}} + c$.
8. $a^2x - \frac{9}{5} a^{4/3} x^{5/3} + \frac{9}{7} a^{2/3} x^{7/3} - \frac{1}{3} x^3 + c$.
9. $\frac{2}{5} x^{5/2} + c$. **10.** $\frac{3}{13} x^{13/3} - \frac{3}{7} x^{7/3} - 6x^{1/3} + c$.
11. $\frac{2x^{2m+\frac{1}{2}}}{4m + 1} + \frac{4x^{m+n+\frac{1}{2}}}{2m + 2n + 1} + \frac{2x^{2n+\frac{1}{2}}}{4n + 1} + c$.
12. $2a^{3/2} x^{1/2} - 4ax + 4a^{1/2} x^{3/2} - 2x^2 + \frac{2}{5} a^{-1/2} x^{5/2} + c$.
13. $\frac{1}{\sqrt{7}} \arctan \frac{x}{\sqrt{7}} + c$. **14.** $\frac{1}{2\sqrt{10}} \log \left| \frac{x - \sqrt{10}}{x + \sqrt{10}} \right| + c$.

15. $\log[x+\sqrt{4+x^2}]+C$. **16.** $\arcsin\frac{x}{2\sqrt{2}}+c$.

17. $\arcsin\frac{x}{\sqrt{2}}-\log[x+\sqrt{(x^2+2)}]+c$.

18. $\tan x-x+c$. **19.** $x-\coth x+c$. **20.** $(3e)^x/(\log 3+1)$. **21.** $b-a$.

22. $v_0T-\frac{1}{2}gT^2$. **23.** 3. **24.** $(2^{10}-1)/\log 2$. **25.** 156. **26.** $\log x$.

27. $-\sqrt{(1+x^4)}$. **28.** $2xe^{-x^4}-e^{-x^2}$. **29.** $\frac{1}{2}x^{-1/2}\cos x+x^{-2}\cos(x^{-1})$.

30. $x=n\pi, n=0, 1, 2, \ldots$ **31.** $\frac{1}{2}$. **32.** $\log 2$. **33.** $(p+1)^{-1}$. **34.** 7/3.

35. 100/3. **36.** 7/4. **37.** 16/3. **38.** $-2/3$. **39.** $\frac{1}{2}\log 2/3$. **40.** $\log 9/8$.

41. $35\frac{1}{15}-32\log 3$. **42.** $\arctan 3-\arctan 2=\arctan 1/7$.

43. $\log 4/3$.

44. $\pi/16$. **45.** $1-1/\sqrt{3}$. **46.** $a\log|c/(a-x)|$. **47.** $x+\log|2x+1|+c$.

48. $-\frac{3}{2}x+\frac{11}{4}\log|3+2x|+c$. **49.** $\frac{x}{b}-\frac{a}{b^2}\log|a+bx|+c$.

50. $\frac{ax}{a}+\frac{ba-a\beta}{a^2}\log|ax+\beta|+c$.

51. $\frac{1}{2}x^2+x+2\log|x-1|$. **52.** $\frac{1}{2}x^2+2x+\log|x+3|+c$.

53. $\frac{1}{4}x^4+\frac{1}{3}x^3+x^2+2x+3\log|x-1|+c$.

54. $a^2x+2ab\log|x-a|-b^2/(x-a)+c$.

55. $\log|x+1|+(x+1)^{-1}$. **56.** $-2b\sqrt{1-x}$. **57.** $-\frac{2}{3b}(a-bx)^{2/3}$.

58. $\sqrt{x^2-4}+3\log|x+\sqrt{x^2-4}|+c$.

59. $\frac{1}{2}\log|x^2-5|+c$. **60.** $\frac{1}{4}\log(2x^2+3)+c$.

61. $\frac{1}{2a}\log(a^2x^2+b^2)+\frac{1}{a}\arctan\frac{ax}{b}+c$. **62.** $\frac{1}{2}\arcsin\frac{x^2}{a^2}+c$.

63. $\frac{1}{3}\arctan x^3+c$. **64.** $\frac{1}{3}\log|x^3+\sqrt{x^6-1}|+c$.

65. $\frac{1}{4}\left(\arctan\frac{1}{2}x\right)^2$. **66.** $\frac{2}{3}(\arcsin x)^{3/2}$. **67.** $(-a/m)e^{-mx}+c$.

68. $C-4^{2-3x}/(3\log 4)$. **69.** $e^x+e^{-x}+c$.

70. $\frac{1}{2}a(e^{2x/a}-e^{-2x/a})+2x+c$.

71. $(\log a-\log b)^{-1}(a^xb^{-x}-a^{-x}b^x)-2x+c$.

72. $\frac{2}{3\log a}a^{3x/2}+\frac{2}{\log a}a^{-\frac{1}{2}x}+c$. **73.** $C-\frac{1}{2}e^{-(1+x^2)}$.

74. $c+\frac{1}{2\log 7}\cdot 7^{x^2}$. **75.** $e^{-1/x}+C$. **76.** $\frac{2}{\log 5}5^{\sqrt{x}}+c$.

77. $\log|e^x-1|+c$. **78.** $C-\frac{2}{3b}(a-be^x)^{3/2}$. **79.** $C-\frac{5}{12}(5-x^2)^{6/5}$.

80. $C+\frac{1}{4}\log|x^4-4x+1|$. **81.** $\frac{1}{4\sqrt{5}}\arctan\frac{x^4}{\sqrt{5}}+c$. **82.** $C-\frac{1}{2}e^{-x^2}$.

83. $\sqrt{\frac{3}{2}}$ arc tan $\left(\sqrt{\frac{3}{2}}x\right) - \frac{1}{\sqrt{3}}$ log $[x\sqrt{3} + \sqrt{2+3x^2}] + c$.
84. $\frac{1}{3}x^3 + x - \frac{1}{2}x^2 - 2\log|x+1| + c$.
85. $C - 2e^{-\frac{1}{2}x}$. **86.** $\log|x + \cos x| + c$.
87. $\frac{1}{3}\left\{\log|\sec 3x + \tan 3x| + \frac{1}{\sin 3x}\right\} + c$.
88. $C - (\log x)^{-1}$. **89.** $\log|\tan x + \sqrt{\tan^2 x - 2}| + c$.
90. $\sqrt{2}$ arc $\tan(x\sqrt{2}) + C - \frac{1}{4(2x^2+1)}$. **91.** $C + a^{\sin x}/\log a$.
92. $\frac{1}{2}(x^3+1)^{2/3} + c$. **93.** $\frac{1}{2}$ arc sin $(x^2) + c$.
94. $\frac{1}{a}\tan(ax) - x + c$.
95. $\frac{1}{2}x - \frac{1}{2}\sin x + c$. **96.** arc sin $\left(\frac{1}{2}\tan x\right) + c$.
97. $a\log\left|\tan\left(\frac{x}{2a} + \frac{1}{4}\pi\right)\right| + c$. **98.** $\frac{3}{4}(1 + \log x)^{4/3} + c$.
99. $C - 2\log|\cos\sqrt{x-1}|$.
100. $\frac{1}{2}\log\left|\tan\left(\frac{1}{2}x^2\right)\right| + c$.
101. $e^{\arctan x} + \frac{1}{4}\log^2(1+x^2) + \arctan x + c$.
102. $C - \log|\sin x + \cos x|$.
103. $C - \sqrt{2}\log\left|\tan\frac{x}{2\sqrt{2}}\right| - 2x - \sqrt{2}\cos\frac{x}{\sqrt{2}} + c$.
104. $x + \frac{1}{\sqrt{2}}\log\frac{x-\sqrt{2}}{x+\sqrt{2}} + c$.
105. $\log|x| + 2\arctan x + C$. **106.** $e^{\sin^2 x} + c$.
107. $\frac{5}{\sqrt{3}}$ arc sin $\frac{x\sqrt{3}}{2} + \sqrt{(4-3x^2)} + c$. **108.** $x - \log(1 + e^x) + c$.
109. $\frac{1}{\sqrt{a^2-b^2}}$ arc tan $\sqrt{\frac{a-b}{a+b}}\,x + c$.
110. $\log\{e^x + \sqrt{e^{2x} - 2}\}$. **111.** $\frac{1}{a}\log|\tan ax| + c$.
112. $C - \frac{T}{2\pi}\cos(2\pi x/T + \Phi_0)$. **113.** $\frac{1}{4}\log|(2 + \log x)/(2 - \log x)| + c$.
114. $C - \frac{1}{2}(\arccos\frac{1}{2}x)^2$.
115. $C - e^{-\tan x}$. **116.** $\frac{1}{2}$ arc sin $\left(\frac{1}{\sqrt{2}}\sin^2 x\right) + c$. **117.** $c - 2\cot 2x$.
118. $\frac{1}{2}(\arcsin x)^2 - \sqrt{1-x^2} + c$.
119. $\log[\sec x + \sqrt{\sec^2 x + 1}] + c$.
120. $\frac{1}{4\sqrt{5}}\log\frac{\sqrt{5} + \sin 2x}{\sqrt{5} - \sin 2x} + c$.
121. $\frac{1}{\sqrt{2}}$ arc tan $\left(\frac{1}{\sqrt{2}}\tan x\right) + c$. **122.** $\frac{2}{3}[\log\{x + \sqrt{x^2+1}\}]^{3/2} + c$.

123. (*a*) $C + 1/\sqrt{2} \text{ arc cos } \sqrt{2}/x$, if $x > \sqrt{2}$; (*b*) $C - \log(1 + e^{-x})$;
(*c*) $C + \frac{1}{80}(5x^2 - 3)^8$; (*d*) $\frac{2}{3}(x + 1)^{3/2} - 2(x + 1)^{1/2} + c$.
(*e*) $\log[\sin x + \sqrt{1 + \sin^2 x}] + c$;

124. $\frac{1}{4}[(2x + 5)^{12}/12 + 5(2x + 5)^{11}/11] + c$.

125. $2\left(\frac{1}{3} x^{3/2} - \frac{1}{2} x + 2\sqrt{x} - 2 \log|1 + \sqrt{x}| + c\right)$:

126. $\log|(\sqrt{2x + 1} - 1)/(\sqrt{2x + 1} + 1)| + c$.

127. $2 \text{ arc tan } \sqrt{e^x - 1}$. **128.** $\log x - \log 2 \log|\log x + 2 \log 2| + c$.

129. $\frac{1}{3}(\text{arc sin } x)^3 + c$. **130.** $\frac{2}{3}(e^x - 2)(e^x + 1)^{1/2} + c$.

131. $\frac{2}{5}(\cos^2 x - 5)\sqrt{\cos x} + c$. **132.** $\log x - \log[1 + \sqrt{x^2 + 1}] + c$.

133. $C - \frac{1}{2} x\sqrt{1 - x^2} + \frac{1}{2} \text{ arc sin } x$.

134. $C - \frac{1}{3} x^2\sqrt{2 - x^2} - \frac{4}{3}\sqrt{2 - x^2} + c$.

135. $\sqrt{x^2 - a^2} - a \text{ arc cos } a/x + c$.

136. $\text{arc cos } 1/x + c$ if $x > 0$ and $\text{arc cos}(-1/x) + c$ if $x < 0$.

137. $\sqrt{x^2 + 1} - \log[1 + \sqrt{x^2 + 1}] - \log x + c$.

138. $C - \frac{1}{4} x^{-1}\sqrt{4 - x^2}$. **139.** $\frac{1}{2} x\sqrt{1 - x^2} + \frac{1}{2} \text{ arc sin } x + c$,

140. $2 \text{ arc sin } \sqrt{x} + c$. **141.** $x \log x - x + c$.

142. $x \text{ arc tan } x - \frac{1}{2}\log(1 + x^2) + c$.

143. $x \text{ arc sin } x + \sqrt{1 - x^2} + c$. **144.** $C + \sin x - x \cos x$.

145. $C + \frac{1}{3} x \sin 3x + \frac{1}{9}\cos 3x$.

146. $C - (x + 1)e^{-x}$. **147.** $C - (x \log 2 + 1)/(2^x \log^2 2)$.

148. $\frac{1}{27}(9x^2 - 6x + 2)e^{3x} + c$. **149.** $C - (x^2 + 5)e^{-x}$.

150. $C - 3e^{-x/3}(x^3 + 9x^2 + 54x + 162)$.

151. $C - \frac{1}{4} x \cos 2x + \frac{1}{8}\sin 2x$.

152. $\frac{1}{4}(2x^2 + 10x + 11)\sin 2x + \frac{1}{4}(2x + 5)\cos 2x + c$.

153. $\frac{1}{3} x^3 \log x - \frac{x^3}{9} + c$. **154.** $x \log^2 x - 2x \log x + 2x + c$.

155. $C - \frac{1}{2} x^{-2}\log x - \frac{1}{4} x^{-2}$. **156.** $2\sqrt{x}\log x - 4\sqrt{x} + c$.

157. $\frac{1}{2}(x^2 + 1) \text{ arc tan } x - \frac{1}{2} x + c$.

158. $\frac{1}{2} x^2 \text{ arc sin } x - \frac{1}{4} \text{ arc sin } x + \frac{1}{4} x\sqrt{1 - x^2} + c$.

159. $x \log[x + \sqrt{1 + x^2}] - \sqrt{1 + x^2} + c$. **160.** $C - x \cot x + \log|\sin x|$.

161. $\frac{1}{2} \text{ arc tan } \frac{1}{2}(x + 1) + c$. **162.** $\frac{1}{2}\log\left|\frac{x}{x + 2}\right| + c$.

163. $\frac{2}{\sqrt{11}} \operatorname{arc\,tan} \frac{6x-1}{\sqrt{11}} + c.$

164. $\frac{1}{2} \log (x^2 - 7x + 13) + \frac{7}{\sqrt{3}} \operatorname{arc\,tan} \frac{2x-7}{\sqrt{3}} + c.$

165. $\frac{3}{2} \log (x^2 - 4x + 5) + 4 \operatorname{arc\,tan}(x - 2).$

166. $x - \frac{5}{2} \log (x^2 + 3x + 4) + \frac{9}{\sqrt{7}} \operatorname{arc\,tan} \frac{2x+3}{\sqrt{7}} + c.$

167. $x + 3 \log (x^2 - 6x + 10) + 8 \operatorname{arc\,tan} (x - 3) + c.$

168. $\frac{1}{\sqrt{2}} \operatorname{arc\,sin} \frac{4x-3}{4} + c.$ **169.** $\operatorname{arc\,sin}(2x - 1) + c.$

170. $\log \left| x + \frac{1}{2} p + \sqrt{x^2 + px + q} \right| + c.$

171. $\frac{1}{a-b} \log \left| \frac{x+b}{x+a} \right| + c.$ **172.** $x + 3 \log |x - 3| - 3 \log |x - 2| + c.$

173. $\frac{1}{12} \log \left| \frac{(x-1)(x+3)^3}{(x+2)^4} \right| + c.$ **174.** $\log \left| \frac{(x-1)^4 (x-4)^5}{(x+3)^7} \right| + c.$

175. $5x + \frac{1}{2} \log x + \frac{161}{6} \log (x - 4) - \frac{7}{3} \log (x - 1) + c.$

176. $(1 + x)^{-1} + \log \left| \frac{x}{x+1} \right| + c.$

177. $\frac{1}{4} x - \frac{7}{16} \log (2x - 1) - \frac{9}{16} \log(2x + 1) + \log x + c.$

178. $\frac{1}{2} x^2 - \frac{11}{(x-2)^2} - \frac{8}{(x-2)} + c.$

179. $C - \frac{9}{2} (x - 3)^{-1} - \frac{1}{2} (x + 1)^{-1}.$

180. $C + \frac{8}{49(x-5)} - \frac{27}{49(x+2)} + \frac{30}{343} \log \left| \frac{x-5}{x+2} \right|.$

181. $C - \frac{1}{2} (x^2 - 3x + 2)^{-2}.$ **182.** $x + \log|x| - \frac{1}{2} \log (x^2 + 1) + c.$

183. $x + \frac{1}{4} \log \left| \frac{x-1}{x+1} \right| - \frac{1}{2} \operatorname{arc\,tan} x + c.$

184. $\frac{1}{52} \log |x - 3| - \frac{1}{20} \log |x - 1| + \frac{1}{65} \log(x^2 + 4x + 5) +$
$+ \frac{7}{130} \operatorname{arc\,tan} (x + 2) + c.$ **185.** $\sin x - \frac{1}{3} \sin^3 x + c.$

186. $C - \cos x + \frac{2}{3} \cos^3 x - \frac{1}{5} \cos^5 x.$ **187.** $\frac{1}{3} \sin^3 x - \frac{1}{5} \sin^5 x + c.$

188. $\frac{1}{4} \cos^8 \frac{x}{2} - \frac{1}{3} \cos^6 \frac{x}{2} + c.$

189. $\frac{1}{2} \sin^2 x - \frac{1}{2} \operatorname{cosec}^2 x - 2 \log|\sin x| + c.$

190. $\frac{3}{8} x - \frac{1}{4} \sin 2x + \frac{1}{32} \sin 4x + c.$ **191.** $\frac{x}{8} - \frac{1}{32} \sin 4x + c.$

192. $\frac{1}{16} x - \frac{1}{64} \sin 4x + \frac{1}{48} \sin^3 2x + c.$

193. $\frac{5}{16} x + \frac{1}{12} \sin 6x + \frac{1}{64} \sin 12x + \frac{1}{144} \sin^3 6x + c.$

194. $C - \cot x - \frac{1}{3} \cot^3 x.$ **195.** $\tan x + \frac{2}{3} \tan^3 x + \frac{1}{5} \tan^5 x + c.$

196. $C - \frac{1}{3}\cot^3 x - \frac{1}{5}\cot^5 x$. **197.** $C + \tan x + \frac{1}{3}\tan^3 x - 2\cot 2x$.

198. $\frac{1}{2}\tan^2 x + 3\log|\tan x| - \frac{3}{2}\cot^2 x - \frac{1}{2}\cot^4 x + c$.

199. $\sec^2 \frac{1}{2}x + 2\log\left|\tan\frac{x}{2}\right| + c$.

200. $\frac{1}{2}\sqrt{2}\left[\log\left|\tan\frac{1}{2}x\right| + \log\left|\tan\left(\frac{1}{2}x + \frac{1}{4}\pi\right)\right|\right] + c$.

201. $C - \frac{1}{4}\cos x \operatorname{cosec}^4 x - \frac{3}{8}\cos x \operatorname{cosec}^2 x + \frac{3}{8}\log\left|\tan\frac{1}{2}x\right| + c$.

202. $\frac{1}{16}\sin 4x \sec^4 4x + \frac{3}{32}\sin 4x \sec^2 4x +$
$+ \frac{3}{32}\log\left|\tan\left(2x + \frac{\pi}{4}\right)\right| + c$. **203.** $\frac{1}{5}\tan 5x - x + c$.

204. $C - \frac{1}{2}\cot^2 x - \log|\sin x|$. **205.** $C - \frac{1}{3}\cot^3 x + \cot x + x$.

206. $\frac{3}{2}\tan^2\frac{x}{3} + \tan^3\frac{x}{3} - 3\tan\frac{x}{3} + 3\log\left|\cos\frac{1}{3}x\right| + x + c$.

207. $\frac{1}{4}x^2 - \frac{1}{8}\sin 2x^2 + c$.

208. $C - \frac{3}{4}\cos^{4/3} x + \frac{3}{5}\cos^{10/3} x - \frac{3}{16}\cos^{16/3} x + c$.

209. $2\sqrt{\tan x} + c$.

210. $\frac{1}{2\sqrt{2}}\log(z^2 + \sqrt{2}z + 1)/(z^2 - \sqrt{2}z + 1) - \frac{1}{\sqrt{2}}\operatorname{arc\,tan}\frac{z\sqrt{2}}{z^2 - 1} +$
$+ c$, where $z = \sqrt{\tan x}$. **211.** $C - \frac{1}{16}\cos 8x + \frac{1}{4}\cos 2x$.

212. $C - \frac{1}{50}\sin 25x + \frac{1}{10}\sin 5x$. **213.** $\frac{3}{5}\sin\frac{5}{6}x + 3\sin\frac{x}{6} + c$.

214. $\frac{3}{2}\cos\frac{x}{3} - \frac{1}{2}\cos x + c$. **215.** $\frac{1}{4a}\sin 2ax + \frac{1}{2}x\cos 2b + c$.

216. $\frac{1}{2}x\cos\varphi - \frac{1}{4\omega}\sin(2\omega x + \varphi)$.

217. $\frac{1}{2}\sin x + \frac{1}{20}\sin 5x + \frac{1}{28}\sin 7x + c$.

218. $\frac{1}{24}\cos 6x - \frac{1}{16}\cos 4x - \frac{1}{8}\cos 2x + c$.

219. $\frac{1}{4}\log\left|\frac{\tan\frac{1}{2}x - 2}{\tan\frac{1}{2}x + 2}\right| + c$.

220. $\frac{1}{\sqrt{2}}\log\left|\tan\left(\frac{1}{2}x + \frac{\pi}{8}\right)\right| + c$. **221.** $x - \tan\frac{1}{2}x + c$.

222. $C - x + \tan x + \sec x$.

223. $\log\left|\frac{\tan\frac{1}{2}x - 5}{\tan\frac{1}{2}x - 3}\right| + c$. **224.** $\operatorname{arc\,tan}(1 + \tan\frac{1}{2}x) + c$.

225. $\frac{12}{13}x - \frac{5}{13}\log|2\sin x + 3\cos x| + c$.

226. $C - \log|\cos x - \sin x|$. **227.** $\frac{1}{2}$ arc tan $\left(\frac{1}{2} \tan x\right)$.
228. $\frac{1}{2}(x + 1)\sqrt{3 - 2x - x^2} + 2 \text{ arc sin } \frac{1}{2}(x + 1) + c$.
229. $\frac{1}{2} x\sqrt{2 + x^2} + \log[x + \sqrt{2 + x^2}] + c$.
230. $\frac{1}{2} x\sqrt{x^2 + 9} - \frac{9}{2} \log[x + \sqrt{x^2 + 9}] + c$.
231. $\frac{1}{2}(x - 1)\sqrt{x^2 - 2x + 2} + \frac{1}{2} \log[x - 1 + \sqrt{x^2 - 2x + 2}]$.
232. $\frac{1}{2} x\sqrt{x^2 - 4} - 2 \log|x + \sqrt{x^2 - 4}| + c$.
233. $C + \frac{1}{4}(2x + 1)\sqrt{x^2 + x} - \frac{1}{8} \log|2x + 1 + 2\sqrt{x^2 + x}|$.
234. 2. **235.** Divergent. **236.** $1/(1 - p)$ if $p < 1$; divergent if $p \geqslant 1$.
237. Divergent. **238.** $\frac{1}{2}\pi$. **239.** Divergent. **240.** 1.
241. $1/(p - 1)$ if $p > 1$; divergent if $p \leqslant 1$.
242. π. **243.** $\pi/\sqrt{5}$. **244.** Divergent. **245.** Divergent. **246.** $1/\log 2$.
247. Divergent. **248.** $1/\log a$. **249.** Divergent. **250.** $1/k$. **251.** $\frac{1}{8}\pi^2$.
252. $\frac{1}{3} + \frac{1}{4} \log 3$. **253.** $\frac{2\pi}{3\sqrt{3}}$. **254.** Divergent. **255.** Convergent.
256. Divergent. **257.** Convergent. **258.** Convergent. **259.** Convergent.
260. Divergent. **261.** Convergent. **264.** 32/3. **265.** 1. **266.** 1/2. **267.** 17/4.
268. 2. **269.** $\log 2$. **270.** $m^2 \log 3$. **271.** πa^2. **272.** 12. **273.** $\frac{4}{3}p^2$. **274.** 9/2.
275. 32/3. **276.** 4. **277.** $\frac{4}{3}a^2$. **278.** 15π. **279.** $\frac{3}{8}\pi ab$. **280.** $3\pi a^2$.
281. $\pi(b^2 + 2ab)$. **282.** $6\pi a^2$. **283.** $\frac{3}{2}a^2$. **284.** $\frac{3}{2}\pi a^2$. **285.** $\frac{1}{8}\pi a^2$.
286. a^2. **287.** $\frac{1}{4}\pi a^2$. **288.** $\frac{9}{2}\pi$. **289.** $\frac{1}{3}(14 - 8\sqrt{2})a^2$.
290. $\pi p^2(1 - e^2)^{-3/2}$.
291. $a^2\left(\frac{1}{3}\pi + \frac{1}{2}\sqrt{3}\right)$. **292.** $\pi\sqrt{2}$. **293.** $\frac{8}{27}(10\sqrt{10} - 1)$.
294. $\sqrt{2} + \log(1 + \sqrt{2})$. **295.** $\sqrt{1 + e^2} - \sqrt{2} - 1 + \log[\sqrt{1 + e^2} - 1] +$
$+ \log(\sqrt{2} + 1)$. **296.** $1 + \frac{1}{2}\log\frac{3}{2}$. **297.** $\log[e + \sqrt{e^2 - 1}]$.
298. $\log(2 + \sqrt{3})$. **299.** $\frac{1}{4}(1 + e^2)$. **300.** $4\sqrt{3}a$. **301.** $\frac{1}{2}aT^2$.
302. $4(a^3 - b^3)/ab$. **303.** $16a$. **304.** $\pi a\sqrt{1 + 4\pi^2} + \frac{1}{2}a\log[2\pi +$
$+ \sqrt{1 + 4\pi^2}]$.
305. $8a$. **306.** $2a[\sqrt{2} + \log(1 + \sqrt{2})]$. **307.** $\pi a^5/30$. **308.** $\frac{4}{3}\pi ab^2$.
309. $\frac{3}{8}\pi^2$. **310.** (a) $\frac{\pi}{4}$; (b) $\frac{4\pi}{7}$. **311.** $\frac{32}{15}\pi a^3$. **312.** $3\pi/10$.
313. $\frac{1}{2}(15 - 16 \log 2)\pi a^3$. **314.** $2\pi^2 a^3$. **315.** $\frac{1}{2}\pi R^2 H$.

316. $\frac{16}{15}\pi h^2 a$. **318.** (a) $5\pi^2 a^3$; (b) $6\pi^3 a^3$; (c) $\frac{1}{6}\pi a^3(9\pi^2 - 16)$.
319. $\frac{32}{105}\pi a^3$. **320.** $\frac{4}{21}\pi a^3$. **321.** $\frac{8}{3}\pi a^3$. **322.** $\pi abh(1 + h^2/3c^2)$.
323. $\frac{4}{3}\pi abc$. **324.** $2\pi[\sqrt{2} + \log(1+\sqrt{2})]$
325. $\pi(\sqrt{5} - \sqrt{2}) + \pi \log 2(\sqrt{2} + 1) - \pi \log(1 + \sqrt{5})$.
326. $2\pi[\sqrt{2} + \log(1 + \sqrt{2})]$.
327. $\frac{1}{4}\pi a^2(e^2 + e^{-2} + 4) = \frac{1}{2}\pi a^2(2 + \sinh 2)$. **328.** $\frac{12}{5}\pi a^2$.
329. $\frac{1}{3}\pi(e - 1)(e^2 + e + 4)$. **330.** $4\pi^2 ab$.
331. (1) $2\pi b^2 + 2\pi ab\,\varepsilon^{-1}\arcsin\varepsilon$;
(2) $2\pi a^2 + \pi b^2 \varepsilon^{-1}[\log(1+\varepsilon) - \log(1-\varepsilon)]$, where $\varepsilon = (a^2 - b^2)^{1/2}/a$.
332. (a) $\frac{64}{3}\pi a^2$; (b) $16\pi^2 a^2$; (c) $\frac{32}{3}\pi a^2$. **333.** $\frac{128}{5}\pi a^2$.
334. $M_x = \frac{1}{2} b\sqrt{a^2 + b^2}$; $M_y = \frac{1}{2} a\sqrt{a^2 + b^2}$.
335. $M_a = \frac{1}{2} ab^2$, $M_b = \frac{1}{2} a^2 b$.
336. $M_x = M_y = \frac{1}{6} a^3$; $\bar{x} = \bar{y} = \frac{1}{3} a$.
337. $M_x = M_y = \frac{3}{5} a^2$; $\bar{x} = \bar{y} = \frac{2}{5} a$. **338.** $2\pi a^2$.
339. $\bar{x} = (a \sin a/a)$, $\bar{y} = 0$. **340.** $\bar{x} = \pi a$, $\bar{y} = \frac{4}{3} a$.
341. $\bar{x} = 4a/3\pi$; $\bar{y} = 4b/3\pi$. **342.** $\bar{x} = \bar{y} = 9/20$. **343.** $\bar{x} = \pi a$, $\bar{y} = \frac{5}{6} a$.
344. $\left(0; 0; \frac{1}{2} a\right)$. **345.** $\left(0; 0; \frac{3}{4} h\right)$. **346.** $\left(0, 0, \frac{3}{8} a\right)$. **347.** πa^3.
348. $I_a = \frac{1}{3} ab^3$, $I_b = \frac{1}{3} a^3 b$. **349.** $\frac{4}{15} hb^3$.
350. $I_a = \frac{1}{4}\pi ab^3$, $I_b = \frac{1}{4}\pi a^3 b$. **351.** $\frac{1}{2}\pi(r_2^4 - r_1^4)$.
352. $\frac{1}{10}\pi r^4 h\,\varrho$ (ϱ = density). **353.** $\frac{2}{5} ma^2$.
354. $V = 2\pi^2 a^2 b$, $S = 4\pi^2 ab$ **355.** $\frac{1}{2}\pi$.

Chapter IV

1. $\frac{1}{2n - 1}$. **2.** $\frac{1}{2n}$. **3.** $n\,2^{1-n}$ **4.** n^{-2}. **5.** $(n + 2)(n + 1)^{-2}$.
6. $2n/(3n + 2)$.
7. $1/n(n + 1)$. **8.** $1.3.5\ldots(2n - 1)/1.4.7\ldots(3n - 2)$. **9.** $(-1)^{n-1}$.
10. $n^{(-1)^{n+1}}$. **11.** Divergent. **12.** Convergent. **13.** Divergent. **14.** Divergent. **15.** Divergent. **16.** Divergent. **17.** Divergent. **18.** Divergent. **19.** Divergent. **20.** Convergent. **21.** Convergent. **22.** Convergent. **23.** Convergent. **24.** Convergent. **25.** Convergent. **26.** Convergent.

27. Convergent. **28.** Convergent. **29.** Divergent. **30.** Divergent.
31. Convergent. **32.** Divergent. **33.** Convergent. **34.** Convergent.
35. Convergent. **36.** Divergent. **37.** Convergent. **38.** Divergent.
39. Convergent. **40.** Divergent. **41.** Divergent. **42.** Convergent.
43. Divergent. **44.** Convergent. **45.** Divergent. **46.** Convergent.
47. Divergent. **48.** Convergent. **49.** Divergent. **50.** Convergent.
51. Convergent. **52.** Convergent. **53.** Divergent. **54.** Divergent.
55. Conditionally convergent. **56.** Conditionally convergent.
57. Absolutely convergent. **58.** Divergent. **59.** Conditionally convergent.
60. Absolutely convergent. **61.** Conditionally convergent.
62. Absolutely convergent. **63.** Absolutely convergent. **64.** Divergent.
65. Absolutely convergent. **66.** Conditionally convergent.
67. Divergent. **68.** Absolutely convergent. **69.** Absolutely convergent.
70. Conditionally convergent. **71.** Divergent. **72.** Absolutely convergent.
73. Absolutely convergent. **75.** Yes. **76.** Convergent.
77. $|R_4| < 1/120$, $|R_5| < 1/720$, $R_4 < 0$, $R_5 > 0$.
78. $R_n < \frac{a_n}{2n+1} + \frac{1}{2^n(2n+1)\,n!}$.
79. $R_n < (n+2)/(n+1)(n+1)!$; $R_{10} < 3\times 10^{-8}$.
80. $-1 < x < 1$. **81.** $-2 \leqslant x < 2$. **82.** $-1 < x < 1$.
83. $-\frac{1}{\sqrt{2}} \leqslant x \leqslant \frac{1}{\sqrt{2}}$. **84.** $-1 < x \leqslant 1$. **85.** $-1 < x < 1$.
86. $-1 < x < 1$. **87.** $-\infty < x < \infty$. **88.** $x = 0$ only.
89. $-\infty < x < \infty$. **90.** $-4 < x < 4$. **91.** $-\frac{1}{3} < x < \frac{1}{3}$.
92. $-2 < x < 2$. **93.** $-\mathrm{e} < x < \mathrm{e}$. **94.** $|z| < 1$. **95.** $|z| < 1$.
96. $|z - 2i| < 3$. **97.** $|z| < \sqrt{2}$. **98.** $z = 0$. **99.** $|z| < \infty$.
100. $|z| < 1/2$. **104.** $-\log(1-x)$, $(-1 \leqslant x < 1)$.
105. $\log(1+x)$, $(-1 < x \leqslant 1)$. **106.** $\frac{1}{2}\log\frac{1+x}{1-x}$, $(|x| < 1)$.
107. $\operatorname{Arc}\tan x$, $(|x| \leqslant 1)$. **108.** $(x-1)^{-2}$, $(|x| < 1)$.
109. $(1-x^2)(1+x^2)^{-2}$, $(|x| < 1)$. **110.** $2(1-x)^{-3}$, $(|x| < 1)$.
111. $x(x-1)^{-2}$, $(|x| > 1)$.
112. $\frac{1}{2}$ [arc tan $x - \frac{1}{2}\log\frac{1-x}{1+x}$, $(|x| < 1)$.
113. $\frac{1}{6}\pi\sqrt{3}$. **114.** 3.
115. $-78 + 59(x+4) - 14(x+4)^2 + (x+4)^3$, $(-\infty < x < \infty)$.
116. $5x^3 - 4x^2 - 3x + 2 + (15x^2 - 8x - 3)\,h +$
$+ (15x - 4)\,h^2 + 5h^3$ $(-\infty < x < \infty,\ -\infty < h < \infty)$.

117. $\sum_{n=1}^{\infty}(-1)^{n-1}(x-1)^n/n$, $(0 < x \leqslant 2)$.

118. $\sum_{n=1}^{\infty}(-1)^n(x-1)^n$, $(0 < x < 2)$.

119. $\sum_{n=0}^{\infty}(n+1)(x+1)^n$, $(-2 < x < 0)$.

120. $e^{-2}\left[1+\sum_{n=1}^{\infty}\frac{(x+2)^n}{n!}\right]$, $(|x| < \infty)$.

121. $2+\sum_{n=1}^{\infty}(-1)^{n-1}\frac{1.3.5\ldots(2n-3)}{4.6.8\ldots 2n}\cdot\frac{(x-4)^n}{2^{2n}}$, $(0 \leqslant x \leqslant 8)$.

122. $\sum_{n=1}^{\infty}(-1)^n\frac{(x-\pi/2)^{2n-1}}{(2n-1)!}$ $(|x| < \infty)$.

123. $\sum_{n=1}^{\infty}(-1)^{n+1}\frac{2^{4n-3}x^{2n}}{(2n)!}$. $(-\infty < x < \infty)$.

124. $1+\sum_{n=2}^{\infty}(-1)^{n-1}\frac{n-1}{n!}x^n$, $(|x| < \infty)$.

125. $8+3\sum_{n=1}^{\infty}\frac{1+2^n+3^{n-1}}{n!}x^n$, $(|x| < \infty)$.

126. $2+\frac{x}{2^2.3.1!}+\sum_{n=2}^{\infty}(-1)^{n-1}\frac{2.5.8\ldots(3n-4)}{2^{3n-1}3^n.n!}x^n$ $(|x| < 8)$.

Chapter V

1. $5/3$; -2. **2.** $(y^2-x^2)/2xy$, $(x^2-y^2)/2xy$, $(y^2-x^2)/2xy$, $2xy/(x^2-y^2)$.
3. $f(x, x^2) = 1+x-x^2$. **4.** $z = R^4/(1-R^2)$. **5.** $f(x) = \sqrt{(1+x^2)}/x$.
6. $\frac{1}{2}(x^2-xy)$. **7.** $f(u) = u^2+2u$; $z = x-1+\sqrt{y}$.
8. $f(y) = \sqrt{1+y^2}$; $z = \sqrt{x^2+y^2}$.
9. (*a*) The interior and circumference of the circle $x^2+y^2 = 1$; (*b*) z only exists along the line $y = x$; (*c*) the half-plane $x+y > 0$; (*d*) The strip $-1 \leqslant y \leqslant 1$; (*e*) The square $|x| \leqslant 1$, $|y| \leqslant 1$; (*f*) $|y| < |x|$; (*g*) $|x| \geqslant 2$, $|y| \leqslant 4$; (*h*) The interior of the annulus bounded by $r = a$, $r = \sqrt{2}a$; (*i*) The regions $2n\pi \leqslant x \leqslant (2n+1)\pi$ in the half-plane $y \geqslant 0$ and the regions $(2n-1)\pi \leqslant x \leqslant 2n\pi$ in the half-plane $y \leqslant 0$; (*j*) $x^2+y > 0$; (*k*) the whole xy-plane; (*l*) the whole xy-plane except $(0, 0)$; (*m*) $x \geqslant 0$, $y > x^{1/4}$; (*n*) the whole xy-plane except the lines $x = 1$ and $y = 0$; (*o*) the interior of the rings $2n\pi \leqslant x^2+y^2 \leqslant (2n+1)\pi$ $(n = 0, 1, 2, \ldots)$.
10. $u_x = yz(xy)^{z-1}$, $u_y = xz(xy)^{z-1}$, $u_z = (xy)^z\log(xy)$.
11. $u_x = yz^{xy}\log z$, $u_y = xz^{xy}\log z$, $u_z = xyz^{xy-1}$.

12. $u_x = 3(x^2 - yz)$, $u_y = 3(y^2 - xz)$, $u_z = 3(z^2 - xy)$.
13. $u_x = 2xy^2\,t/(1 - z^2)$, $u_y = 2x^2\,yt/(1 - z^2)$, $u_z = 2x^2\,y^2\,zt/(1 - z^2)^2$, $u_t = x^2\,y^2/(1 - z^2)$.
14. $u_x = -xr^{-3}$, $u_y = -yr^{-3}$, $u_z = -zr^{-3}$.
15. $u_x = y^2z/(x + y)^2(z + 1)$, $u_y = x^2z/(x + y)^2(z + 1)$, $u_z = xy/(x + y)(z + 1)^2$.
16. $f_x = 1$, $f_y = 1/2$, $f_z = 1/2$. **17.** $r^2 \sin\theta\, dr d\theta\, d\varphi$.
28. $\partial z/\partial x = -y/(x^2 + y^2)$, $dz/dx = 1/(1 + x^2)$.
29. $\partial z/\partial x = yx^{y-1}$, $dz/dx = x^y\left[\varphi'(x)\log x + \frac{y}{x}\right]$.
30. $\partial z/\partial x = 2x\,\partial f/\partial u + ye^{xy}\,\partial f/\partial v$, $\partial z/\partial y = -2y\,\partial f/\partial u + xe^{xy}\,\partial f/\partial v$.
31. $\partial z/\partial u = 0$, $\partial z/\partial v = 1$.
32. $\partial z/\partial x = y\left(1 - \frac{1}{x^2}\right) f'(u)$, $\partial z/\partial y = \left(x + \frac{1}{x}\right) f'(u)$, $\left(u = xy + \frac{y}{x}\right)$.
36. $\partial^2 z/\partial x^2 = abcy^2/(b^2x^2 + a^2y^2)^{3/2}$, $\partial^2 z/\partial x\,\partial y = -abcxy/(b^2x^2 + a^2y^2)^{3/2}$, $\partial^2 z/\partial y^2 = abcx^2/(b^2x^2 + a^2y^2)^{3/2}$.
37. $\partial^2 z/\partial x^2 = 2(y - x^2)/(x^2 + y)^2$, $\partial^2 z/\partial x\,\partial y = -2x/(x^2 + y)^2$, $\partial^2 z/\partial y^2 = -1/(x^2 + y)^2$.
38. $\partial^2 z/\partial x\,\partial y = xy/(2xy + y^2)^{3/2}$.
44. $z = 0$ (min.) at $(1, 0)$. **45.** No maximum or minimum.
46. $z = -1$ (min.) at $(1, 0)$. **47.** $z = 108$ (max.) at $(3, 2)$.
48. $z = -8$ (min.) at $(\sqrt{2}, -\sqrt{2})$ and at $(-\sqrt{2}, \sqrt{2})$.
49. $z = \dfrac{ab}{3\sqrt{3}}$ (max.) at. $(\pm a/\sqrt{3}, \pm b/\sqrt{3})$

$z = -\dfrac{ab}{3\sqrt{3}}$ (min.) at $(\pm a/\sqrt{3}, \mp b/\sqrt{3})$.

50. $z = 1$ (max.) at $(0, 0)$. **51.** $z = 0$ (min.) at $(0, 0)$.
52. $z = \sqrt{3}$ (max.) at $(1, -1)$.
53. (*a*) The two functions $z = 3 \pm \sqrt{25 - (x - 1)^2 - (y + 2)^2}$ are defined by the equation. The $(+)$ sign gives a function which has a maximum value 8 at the point $(1, -2)$; this function takes its minimum value 3 at all points on the circumference of the circle $(x - 1)^2 + (y + 2)^2 = 25$, the function not being defined at points outside this circle. The $(-)$ sign gives a function which has a minimum value -2 at $(1, -2)$; this function takes its maximum value 3 at points on $(x - 1)^2 + (y + 2)^2 = 25$, not being defined at points outside the circle.
(*b*) One function has a maximum $(z = -2)$ at $(-1, 2)$ while the other has a minimum $(z = 1)$ at $(-1, 2)$; both functions reach their other bounding value at points on the curve $4x^3 - 4y^2 - 12x + 16y - 33 = 0$.

54. z has max. value $1/4$ at $(1/2, 1/2)$.

55. z has max. value 5 at $(1, 2)$ and min. value -5 at $(-1, -2)$.

56. z has min. value $36/13$ at $(18/13, 12/13)$.

57. z has max. value $1 + \frac{1}{2}\sqrt{2}$ at $\left(\frac{7\pi}{8} + k\pi, \frac{9\pi}{8} + k\pi\right)$,
and min. value $1 - \frac{1}{2}\sqrt{2}$ at $\left(\frac{3\pi}{8} + k\pi, \frac{5\pi}{8} + k\pi\right)$.

58. u has min. value -9 at $(-1, 2, -2)$ and max. value 9 at $(1, -2, 2)$.

59. u has maximum value a^2 at $(\pm a, 0, 0)$ and minimum value c^2 at $(0, 0, \pm c)$.

60. The cube with edge $V^{1/3}$.

61. The box has a square base of edge $(2V)^{1/3}$ and has height $\left(\frac{1}{4}V\right)^{1/3}$.

62. An equilateral triangle of side $\frac{2}{3}p$. **63.** Cube.

64. The triangle with sides $3p/4$, $3p/4$, $p/2$ rotated about the smallest side.

65. P is at the centre of mass $\left(\frac{m_1x_1 + m_2x_2 + m_3x_3}{m_1 + m_2 + m_3}, \frac{m_1y_1 + m_2y_2 + m_3y_3}{m_1 + m_2 + m_3}\right)$.

66. $\frac{x}{a} + \frac{y}{b} + \frac{z}{c} = 3$.

67. If the ellipsoid has semi-axes a, b, c the rectangular parallelepiped should have edges $2a/\sqrt{3}$, $2b/\sqrt{3}$, $2c/\sqrt{3}$.

INDEX